TREE THINKING

An Introduction to Phylogenetic Biology

David A. Baum and
Stacey D. Smith

ROBERTS AND COMPANY PUBLISHERS
Greenwood Village, Colorado

Tree Thinking: An Introduction to Phylogenetic Biology

Roberts and Company Publishers, Inc.
4950 South Yosemite Street, F2 #197
Greenwood Village, CO 80111 USA
Tel: (303) 221-3325
Fax: (303) 221-3326
Web: www.roberts-publishers.com
Email: info@roberts-publishers.com

Publisher: Ben Roberts
Production editor: Susan Riley at Side By Side Studios
Designer: Mark Ong at Side by Side Studios
Artist: Laurel Muller, cohographics
Copyeditor: Michael Zierler
Proofreader: Jennifer McClain

Front cover:
Klimt, Gustav (1862-1918).
Sketches for the Frieze for the Palais Stoclet in Brussels, Belgium. Watercolour and pencil.
Museum fuer Angewandte Kunst, Vienna, Austria
Photo Credit: Erich Lessing / Art Resource, NY

© 2013 by Roberts and Company Publishers, Inc.
Reproduction or translation of any part of this work beyond that permitted by Section 107 or 108 of the 1976 United States Copyright Act without permission of the copyright owner is unlawful. Requests for permission or further information should be addressed to the Permissions Department at Roberts and Company Publishers.

ISBN: 9781936221165

Library of Congress Cataloging-in-Publication Data

Baum, David A.
Tree thinking : an introduction to phylogenetic biology / David A. Baum and Stacey D. Smith.
p. cm.
Includes index.
ISBN 978-1-936221-16-5
1. Phylogeny. I. Smith, Stacey D. II. Title.
QH367.5.B38 2012
576.8'8--dc23
2012013220

10 9 8 7 6 5 4 3 2 1

For all of our teachers and students

Contents

Preface xv
About the Authors xix

PART I: INTRODUCTION

1 Phylogenetic Trees and Their Importance in Modern Biology 1

PART II: INTERPRETING TREES

2 Tree Thinking and Its Importance in the Development of Evolutionary Thought 9

3 What a Phylogenetic Tree Represents 35

4 Trait Evolution 77

5 Relatedness and Taxonomy 107

6 Gene Trees and Species Trees 139

PART III: INFERRING TREES

7 Phylogenetic Inference with Parsimony 173

8 Phylogenetic Inference with Distance, Maximum Likelihood, and Bayesian Methods 217

9 Statistical Tests of Phylogenetic Hypotheses 265

PART IV: USING TREES

10 Using Trees to Study Character Evolution 305

11 Using Trees to Study Space, Time, and Evolutionary Diversification 349

PART V: RESOURCES

Appendix 1: Data for Phylogenetic Analysis 389
Appendix 2: Generating a Morphological Data Matrix 399
Answers to Chapter Quizzes 405
References 425
Glossary 439
Index 459

Detailed Contents

Preface xv
About the Authors xix

PART I: INTRODUCTION

1 Phylogenetic Trees and Their Importance in Modern Biology 1
The importance of phylogenetic trees 2
From individuals to populations to species 2
Visualizing and modeling how traits evolve 3
Understanding what it means to be related 4
Organizing knowledge of biological diversity 4
Reconstructing history 5
Tree thinking and biological literacy 7

PART II: INTERPRETING TREES

2 Tree Thinking and Its Importance in the Development of Evolutionary Thought 9
Before Charles Darwin 9
Charles Darwin's role in the development of tree thinking 12
Evidence for evolution and common ancestry 15
The emergence of phylogenetic systematics 23
Vestiges of ladder thinking in the popular understanding of evolution 28

3 What a Phylogenetic Tree Represents 35
The continuity of reproduction from the population to the tree of life 35
Asexual organisms 42

Tree terminology and conventions 43
Tree topology 45
Different tree styles 48
Merging and pruning 49
The time axis 53
Branch length and evolutionary rate 55
Communicating phylogenetic uncertainty 58
Unrooted trees 61
Tree-to-tree distances 65

4 Trait Evolution 77
Trait evolution in a single lineage 77
Trait evolution in a branching tree 80
Ancestral and derived character states 84
The evolution of DNA sequences 86
An introduction to homology 90
Homoplasy and consistency 93
Parsimony as a way to infer the history of traits 95

5 Relatedness and Taxonomy 107
The concept of relatedness 107
Relatedness and phylogenetic trees 110
Taxonomy and phylogeny 117
Monophyly and phylogenetic systematics 120
Varieties of non-monophyly 124
Converting monophyletic taxonomies into trees 126
Taxonomic ranks 128
Phylogenetic nomenclature 131

6 Gene Trees and Species Trees 139
Gene trees in sexual populations 140
Gene trees in branching population trees 146
Gene duplication 151
Population reticulation 157

Tree thinking given reticulation 160
The concept of species 161

PART III: INFERRING TREES

7 Phylogenetic Inference with Parsimony 173
A biological example: Carnivora 173
Hennigian inference 178
The maximum parsimony criterion 181
A justification of parsimony 185
Finding optimal trees 187
Parsimony analysis of the carnivoran morphological data 192
Parsimony analysis of DNA sequence data 195
Character and character state weighting 202
Problems of parsimony 205

8 Phylogenetic Inference with Distance, Maximum Likelihood, and Bayesian Methods 217
Introduction to modeling molecular evolution 217
More realistic models of molecular evolution 225
Distance methods 231
Maximum likelihood 238
Choosing an evolutionary model 246
Bayesian inference 247

9 Statistical Tests of Phylogenetic Hypotheses 265
Detecting nonrandomness in a data set 266
Measures of clade support 271
Nonparametric bootstrap analysis 273
Parametric bootstrapping 281
Bayesian statistical analysis 284
Comparing data sets 285
Partition homogeneity 288
Species tree methods 292

PART IV: USING TREES

10 Using Trees to Study Character Evolution 305
Historical chronicles and narratives 305
Testing hypotheses of adaptation 306
The comparative method 311
Model-based analysis of the evolution of discrete traits 316
Evolutionary trends 320
Model-based tests of correlated evolution 323
Stochastic mapping 325
The evolution of continuous traits 327
Detecting correlated evolution of continuous traits 336

11 Using Trees to Study Space, Time, and Evolutionary Diversification 349
Reconstructing dispersal history 349
Vicariance 352
Coevolution and cophylogenetics 355
Phylogenetic community ecology 359
Principles of molecular dating 362
The molecular clock model 363
Relaxed clock models 366
Calibrating a chronogram 369
Clade size and diversification rate 370
Key innovations 375
Joint estimation of character evolution and diversification 378

PART V: RESOURCES

Appendix 1: Data for Phylogenetic Analysis 389
Molecular sequences 389
Molecular presence/absence data 391
Molecular structural data 392
Morphology and other morphotypic data 395
Frequency data 396
Distance data 397

Appendix 2: Generating a Morphological Data Matrix 399
Answers to Chapter Quizzes 405
References 425
Glossary 439
Index 459

Preface

This book is intended to introduce the discipline of phylogenetic biology to advanced undergraduate students, teachers, and biologists in allied disciplines (e.g., biochemistry, cell biology, physiology, ecology). Phylogenetic biology is of great and growing practical significance for answering such diverse questions as, How did this gene acquire its function? Where did this disease originate? and Why does this species exist in this ecological community? Although modern methods for addressing such questions often involve sophisticated computational approaches, detailing these technical aspects of phylogenetics is not what really motivated this book. Rather, as indicated by the book's title, our deeper mission is to help you improve your ability to think clearly and accurately about evolutionary history through the use of phylogenetic trees.

Tree thinking is not just a practical skill to be learned, like riding a bicycle or doing long multiplication. It certainly is an important tool for the biologist, but it offers much more than that. The evolutionary perspective offered by tree thinking helps one appreciate biological diversity in much the same way that some knowledge of music theory can help one appreciate a great symphony. Trees provide you with a conceptual device for exploring the patterns seen in nature and often stimulate hypotheses about the processes underlying these patterns. We wanted to write this book to help readers develop a deeper appreciation for evolutionary theory and how it helps us make sense of both the deep unity and the bewildering diversity of life on earth.

We recognize that this is a very ambitious goal. Changing ways of thinking is the greatest challenge for the educator. And tree thinking is particularly difficult. For a start, it has a significant visual and spatial element that can be hard to transmit through the flat pages of a textbook. Also, tree thinking runs counter to standard perceptions of evolution in popular culture. We do not know why it should be so, but we have learned from working with thousands of students that, without contrary training, people tend to have a one-dimensional and progressive view of evolution. We tend to tell evolution as a story with a beginning, a middle, and an end. Against that backdrop, phylogenetic trees are

challenging; they are not linear but branching and fractal, with one beginning and many equally valid ends. Tree thinking is, in short, counterintuitive.

To achieve our goal of helping you develop as tree thinkers, we complement the text with numerous figures and deploy vivid metaphors to drum home the visual aspects of tree thinking. We also try to engage the reader by including challenging quizzes at the end of each chapter that are designed to help uncover common preconceptions. Finally, whenever possible, we draw on biological examples that we hope you will find interesting.

We should be clear: our emphasis on concepts comes at a cost. This book is not a how-to guide for phylogenetic analysis. We do not provide readers with advice on how to run phylogenetics computer programs, nor even with an exhaustive list of the software packages available. This book is better seen as a why-to guide, but one that nonetheless tries to explain the underlying logic of many of the most commonly deployed methods. Our hope is that you will finish the book not as an expert in phylogenetic analysis but enthused about phylogenetic biology, eager to know more, and equipped with a solid conceptual foundation upon which to build further understanding.

Like any book, this one has been a major endeavor in which many people have played a part. We are very grateful to three former Baum lab students, Ivalú Cacho, Rebecca (Oldham) Haltom, and Margaret (Koopman) Hanes, for their help in hatching the idea of a book on tree thinking and getting us started.

We thank the following for their valuable reviews of drafts: Brent Berger, Sydney Cameron, Donna Daniels, Christopher Dietrich, Luke Harmon, Allan Larson, Stephanie Lyon, Susan Masta, Nicholas Matzke, Mara McDonald, Louise Mead, Susan Offner, Richard Olmstead, Rhiannon Peery, Andrew Petto, Richard Ree, Michael Sanderson, William Saucier, Emily Sessa, David Spooner, Millard Sussman, Jana Vamosi, George Weiblen, James Whitfield, and Barbara Whitlock. Additionally, feedback was provided by the University of Wisconsin Zoology/Evolution Reading Group, the University of Illinois Cameron-Dietrich-Soto-Whitfield lab group, and the University of Connecticut EEB Spring 2008 Systematics Seminar.

Expert advice and information was generously provided by David Ackerly, Cécile Ané, Jeannine Cavender-Bares, Yujin Chung, Julia Clarke, Bret Larget, Allan Larson, Paul Lewis, Brian O'Meara, Sarah Otto, Michael Sanderson, Elliott Sober, David Swofford, Jeffrey Thorne, and George Weiblen.

We would like to thank the members of our labs for their help in this project: Rachel Coburn, Raul Correa, Daniel Gates, Talline Martins, Abigail Mazie,

Nathan Muchhala, Latifa Obaidi, C. R. Pulikesi, Alison Scott, and John Stanga. We are also very grateful for the support and patience of our families: Gabriela Baum, Leora Baum, Micah Baum, Ruth Litovsky, Kathleen R. Smith, and Roger A. Smith, III.

We would like to thank Ben Roberts, our publisher, for his tireless enthusiasm and endless patience, Michael Zierler, our copyeditor, and Jennifer McClain, our proofreader, for important corrections and useful edits, artist Laurel Muller for high-quality work, Mark Ong, design and production manager and art director, for guiding us so expertly through the process of publishing a book, and Susan Riley for composition and layout.

This project was greatly aided by sabbatical support to DB from the John Simon Guggenheim Memorial Foundation, the University of Wisconsin College of Letters and Science, and the Master and Fellows of St. Catherine's College, Oxford.

David A. Baum
Madison, Wisconsin

Stacey D. Smith
Lincoln, Nebraska

February 1, 2012

About the Authors

DAVID A. BAUM

David A. Baum is a professor at the University of Wisconsin, where he conducts research in plant systematics, systematic theory, and plant evolutionary developmental genetics. He earned his PhD in population and evolutionary biology from Washington University in 1991, conducted postdoctoral research at the University of Wisconsin, and then served as assistant and associate professor of organismic and evolutionary biology at Harvard University. He returned to the University of Wisconsin in 2001, and is now professor and chair of botany and director of the James F. Crow Institute for the Study of Evolution.

Baum's research has earned him awards, including an Alfred P. Sloan Foundation Young Investigator Award, an NSF Career Award, and a John Simon Guggenheim Foundation fellowship. In 2006 he was elected Fellow of the American Association for the Advancement of Science for "outstanding contributions in the investigation of plant evolution and for excellence in teaching and mentoring in phylogenetics and evolutionary biology." Baum teaches phylogenetic biology at both the introductory and advanced levels and has been offering workshops on phylogenies and tree thinking for K–12 teachers since 1999.

STACEY D. SMITH

Stacey D. Smith is professor at the University of Nebraska-Lincoln. Her research focuses on the evolution of floral diversity and spans the fields of phylogenetics, evolutionary genetics, comparative methods, and pollination ecology. Supported by a British Marshall fellowship, she earned an MPhil in botanical diversity from the Universities of Reading and Birmingham in the United Kingdom in 2001. She returned to the United States to pursue a doctoral degree in botany at the University of Wisconsin-Madison. After obtaining her PhD in 2006, she conducted postdoctoral research at Duke University through a National Institutes of Health Ruth L. Kirschstein fellowship before joining the faculty at the University of Nebraska in 2010.

Smith teaches introductory organismal biology and phylogenetic biology for undergraduate and graduate students and sponsors outreach events to promote public understanding of plant biology and evolution.

CHAPTER ONE

Phylogenetic Trees and Their Importance in Modern Biology

Open any book or article on the evolution of life on earth and there is a good chance that you will find a depiction of evolution in a treelike form. These evolutionary trees, or phylogenies, represent the diversification of species over time. This book is about phylogenies: what they are, how they are constructed, and what they can tell us.

The purpose of this book is twofold. First, we aim to help readers develop tree-thinking skills. ***Tree thinking*** is the ability to visualize evolution in tree form and to use tree diagrams to communicate and analyze evolutionary phenomena. Tree thinking is essential for developing an accurate understanding of evolution and also helps one to organize knowledge of biological diversity. Second, we provide an introduction to the technical aspects of phylogenetics. We provide enough detail that the rigor and elegance of this field should become clear, without getting bogged down in technical minutiae. If phylogenetics is a car, we aim to teach you how to drive it (use tree thinking to organize evolutionary knowledge) and to show you what is going on inside the engine (how the methods work in principle). However, we do not aspire to teach you to be mechanics. For that we refer you to more advanced books, such as Felsenstein's *Inferring Phylogenies* (2004).

In this introductory chapter, we would like to give you a sense of the profound impact and broad reach of tree thinking in both pure and applied biology. We will also articulate the overall case for tree thinking as an essential element of biological literacy.

This book has a website where you will find a pretest that you can use to assess your current grasp of phylogenetic trees. This test and the tests provided at the end of all subsequent chapters should help you to gauge your own

progress in developing tree-thinking skills and putting them to work to make sense of evolutionary history.

THE IMPORTANCE OF PHYLOGENETIC TREES

With regard to evolution, the terms "tree," "phylogenetic tree," and "phylogeny" are used interchangeably to refer to branching evolutionary histories or to graphs that represent these evolutionary histories. Phylogenetic trees play a prominent role in modern biology because they provide a concise way to visualize evolution as descent from common ancestors.

It is hard to think accurately about evolution and the biological products of evolution (organisms, genes, traits) without reference to a branched structure that loosely resembles a living tree with branches that diverge but do not grow back together. Other models for thinking about and studying evolution, for example, the ladder of life or the Great Chain of Being, are either ineffective or misleading. Admittedly, a tree model is an oversimplification that does not perfectly mirror reality. However, there is strong evidence from many kinds of data (DNA, fossils, physical traits) that the evolutionary history of species (and even more so, of genes) nearly always involves the branching of evolutionary lineages without subsequent fusion. Evolution really seems to be treelike.

Given the treelike form of evolutionary history, tree thinking is a critical skill for modern biologists. Tree thinking is required across many disciplines: from molecular biology to genetics, developmental biology to ecology. Furthermore, tree thinking is important in applied research, such as tracking diseases, documenting responses to climate change, and guiding conservation policies.

FROM INDIVIDUALS TO POPULATIONS TO SPECIES

Phylogenetic trees provide a natural way to understand the continuum of life from individuals to species. Every living organism on the planet has either one or two direct ancestors, and these ancestors likewise had ancestors. Thus, we can imagine retracing the evolutionary history of any living organism by walking backward through time, from ancestor to ancestor.

A key insight of evolutionary theory is that if we start such an "ancestry walk" from two different living organisms, their paths will eventually converge:

we will eventually come upon organisms, ***common ancestors***, which are ancestral to both of the living organisms we started with. How long we have to walk before finding common ancestors will vary depending on our choice of starting organisms, but in all cases we will eventually find them. If we started with two humans, we would not have to go very far back in time to find a common ancestor. If we started with a human and a chimpanzee we would have to go further back in time, and further still if we started with a human and a lily.

Common ancestry provides a natural way to understand the connections between individuals of different populations and different species, and tree diagrams are a natural way to visualize and summarize common ancestry. The tree structure works well at diverse scales, from the inheritance of single genes within a single population, to the relationships of closely related species, to the evolutionary links among lineages as diverse as plants, animals, and bacteria. We explore the connections between these levels in detail in Chapter 3.

VISUALIZING AND MODELING HOW TRAITS EVOLVE

A central concept in evolution is that traits change over time. Phylogenetic trees allow us to visualize the history of trait evolution and to discover general patterns. We can imagine how a certain trait, arising along a particular branch of the tree of life, might come to characterize all organisms descended from that branch. By arranging living and fossil taxa on a model of the tree of life, we can learn about the order in which traits arose or were lost. When we do this, certain patterns become obvious. Traits that evolved recently, such as the evolution of spoken language in human ancestors, are only found in one or a few living species. In contrast, traits that evolved longer ago, like having a backbone, are often (but not always) present in many living species. Such patterns are explicable given the treelike form of evolutionary history.

Trees also allow us to reconstruct the evolutionary changes that have occurred in any kind of trait that is passed down from one generation to the next: from physical traits like tail length or eye color to geographic distribution and social behavior. Importantly, we can distinguish homologous traits from analogous traits (Chapter 4). For example, mammary glands in mammals are similar across species because they descend from a common ancestor, but the wings of birds, bats, and insects evolved independently and are similar because they represent adaptations for similar biological functions.

UNDERSTANDING WHAT IT MEANS TO BE RELATED

We commonly see statements in the media like, "humans and chimps are each other's closest relatives" or "recently discovered fossil dinosaur is closely related to birds." But what does it mean to be closely or distantly related? Chapter 5 explains that the degree of relatedness of two organisms is defined by how far back we must go to reach their most recent common ancestor. In this book, we hope to equip readers with the skills to make statements about relatedness by reading phylogenetic trees.

Defining relatedness in terms of common ancestry allows us to create evolutionarily meaningful taxonomic systems. In such systems, organisms are assigned to named groups, called ***taxa*** (singular = ***taxon***), that align with evolutionary history: all the organisms in a taxon are more closely related to each other than to any organism outside the taxon. As discussed in Chapter 5, taxa are easily delimited when we know the phylogenetic tree for the organisms in question.

Having a taxonomic system that reflects evolutionary history is important because it has predictive value. If a taxonomic system mirrors the actual evolutionary history, members of a taxon are likely to have many traits in common. For example, once we know that a newly discovered animal is a mammal, we can make numerous predictions about its physiology and anatomy—and most of those predictions will be true. With a taxonomic system that does not fully reflect evolutionary history, predictions will be less reliable.

ORGANIZING KNOWLEDGE OF BIOLOGICAL DIVERSITY

Trees provide an excellent framework for efficiently managing biological information that varies across species and/or across time. Once you have defined the outline of a tree, you can annotate it by attaching to its branches lists of features that are shared by all descendants of that branch. This provides a way to keep track of which organisms have which traits. For example, once you learn the basic structure of vertebrate phylogeny, you can readily ornament it with traits such as lungs, hair, feathers, and endothermy. The tree is a much more efficient way to store such information than trying to recall the full list of traits that occur in a particular organism. As will be shown more formally in Chapter

4, the advantage of using a tree for storing information about the distribution of traits grows rapidly as the number of taxa and traits increases.

Trees can depict not only the histories of species lineages and organismal traits but also the histories of genes. New genes generally arise by the duplication of an existing gene (Chapter 6), followed by divergence of one of the copies of the duplicated gene. In fact, it has been argued that all genes in organisms today evolved by duplication and divergence from a small number of genes present in the universal ancestor. In the same way that species can be grouped into taxa at different ranks, individual genes can be assigned to gene families and subfamilies. The position of a gene in a gene tree can tell us a lot about its likely properties. For example, the genes that control plant pigment production all fall into a family of genes with similar biochemical functions. Consequently, gene trees have become important tools in molecular biology. Open any scientific journal in molecular biology or genetics and you will see that many or most of the research papers that describe newly isolated genes include phylogenetic results that attempt to place the gene into a family or subfamily.

RECONSTRUCTING HISTORY

There are many applications for phylogenetic trees. Sometimes the motivation is just to understand the natural world and how it came to be: When did language evolve? Why were the flowering plants so successful? What colors could *Tyrannosaurus rex* see? Other questions have more direct practical implications: Where did invasive zebra mussels in the Great Lakes come from? From what animal source did humans become infected by swine flu?

You may wonder, How can evolutionary trees make it possible to answer such questions? While the answer will be more fully addressed in subsequent chapters, we will mention general considerations here. First, by knowing something about the rate at which evolution occurs and by attaching fossils to parts of the tree, we can convert abstract branch lengths into units of real time (Chapter 11). This allows us to make statements like "humans and flies diverged from a common ancestor about 600 million years ago (Ma)" or "the ancestors of birds evolved feathers about 150 Ma." Furthermore, these "time-calibrated" trees can be combined with geographical information to learn where these events happened. Recent studies have suggested that New World monkeys, such as howler and spider monkeys, arose from a common ancestor

that dispersed across the Atlantic to South America from Africa (Schrago and Russo 2003). As improbable as it may seem that a few monkeys floated across the Atlantic on a raft of vegetation, the recency of the split between African and South American monkeys, around 35 Ma, makes the alternative hypothesis, that monkeys walked over before the continents separated, some 100 Ma, even less likely.

Using phylogenies to understand spatial and temporal patterns is particularly powerful when it comes to tracking human pests and invasive species. One excellent example is the case of severe acute respiratory syndrome (SARS), a previously unknown viral disease that emerged in Asia in 2002 and then spread rapidly around the world. Phylogenetic analysis of viral DNA from infected patients showed that the causative agent of SARS was a coronavirus, a group of viruses already known to infect the respiratory and digestive systems of mammals and birds. Because the virus first appeared in Asia, the simplest explanation would be that it was derived from a strain of SARS that was present in other animals in Asia. Subsequent sampling of the DNA from a wide array of animals (e.g., bats, cows, and civets—the latter being small carnivores) showed that SARS probably originated in bats and subsequently infected humans and civets (Lau et al. 2005). Since civets are a food item in parts of Asia, the frequent contact between humans and civets may have promoted the exchange and spread of the virus. There was also one case in which a SARS virus found in a cow was most closely related to a SARS virus found in a human. This might be an instance in which the SARS virus hopped from humans to cows. Similar sorts of analyses have also been used to solve other epidemiological and forensic mysteries, such as the now-famous case of the Florida dentist who passed HIV to six of his patients (Hillis and Huelsenbeck 1994).

Beyond tracking the evolution of single traits, simultaneous analysis of multiple traits allows us to address other kinds of questions. For example, phylogenetic methods can help us reconstruct the intermediate steps by which characters composed of seemingly interdependent parts could have evolved. One such character is powered flight in birds. Flying depends on a large number of functionally interrelated features, including feathers, hollow bones, and elongated forearms. Since it might seem that none of these traits would be very useful on its own, it is not clear how flight could have evolved. However, phylogenetic studies of extinct dinosaurs have documented the stepwise origin of many of these flight-related traits and have shed light on their possible roles prior to

the evolution of flight. This work has greatly advanced our understanding of how birds took to the air.

Another use of phylogenetic trees involves the study of statistical patterns in the distribution of traits across species. As outlined in Chapter 10, there are many ways that phylogenetic trees can aid in the study of trait evolution. For example, suppose you suspect that red flowers evolved in diverse plant groups through selection for the attraction of hummingbird pollinators. Such a hypothesis could be tested by using phylogenetic trees to reconstruct the evolutionary history of both flower color and pollination type. Then you could see if switches in flower color followed shifts toward hummingbird pollination more often than you would expect by chance, a result which would support the hypothesis that hummingbirds exert selection on flower color.

TREE THINKING AND BIOLOGICAL LITERACY

While the importance of tree thinking has been apparent to many biologists since at least the 1980's, there has been no easy way for students to learn tree thinking. This book aims to fill this vacuum by providing a thorough but accessible introduction to tree thinking and phylogenetics.

The most basic aspects of tree thinking (which are covered in the next five chapters) involve understanding what phylogenetic trees are, and being able to use them to think clearly about evolution. At a practical level, tree thinking allows you to correctly extract information from tree diagrams. Tree thinking also entails an ability to convert verbal or textual discussions of evolutionary relationships or classifications into phylogenies. The latter skill is critical so that you can begin building an internal representation of the tree of life as a basis for organizing knowledge of biological diversity. This helps overcome the tendency to treat biology as an overwhelming body of disconnected information.

Trees discussed by biologists are the result of scientific inference. Nobody directly "sees" a tree any more than they directly "see" the Krebs cycle or the base pairing in a DNA molecule. Therefore, a deeper level of tree thinking requires understanding the principles that allow scientists to infer phylogenies from biological data. Additionally, we have found in our teaching that the ability to think clearly about trees can be aided by working through the methods that are used to infer trees. Therefore, Chapters 7–9 explore the procedures for inferring phylogenies, emphasizing the underlying concepts and principles.

While this section is not intended to serve as a how-to guide, it will introduce the mathematical principles and provide pointers to other, more technical resources.

The deepest level of tree thinking involves not just understanding trees and where they come from, but also knowing how to put trees to use. The last two chapters of the book explore some of the many ways in which biologists put trees to work. These chapters survey the principles and analytical methods by which trees are used to elucidate various aspects of recent and ancient evolutionary history.

CHAPTER TWO

Tree Thinking and Its Importance in the Development of Evolutionary Thought

The origin of a scientific understanding of evolution occurred primarily through the work of Charles Darwin and Alfred Russel Wallace in the mid-nineteenth century. Their theories depended heavily on the concept of common ancestry and, thus, implicitly demanded tree thinking. In this chapter we explore the historical transition from ladder thinking to tree thinking. ***Ladder thinking*** is the idea that living species represent a continuum of forms of differing degrees of advancement, with humans near the "top" of a metaphorical ladder. We will explain why tree thinking can be reconciled with the Darwinian notion of descent from common ancestry, but ladder thinking cannot. We also describe how evolutionary trees and common ancestry have provided a framework for identifying evidence for evolution. Finally, we show that despite the importance of trees within evolutionary biology, vestiges of ladder thinking are common in verbal and visual depictions of evolution. We argue that it is critical to expunge ladder thinking completely and replace it with clear tree thinking if the broader scientific and lay community is to arrive at an accurate understanding of evolution.

BEFORE CHARLES DARWIN

Jean Baptiste de Lamarck (1744–1829), naturalist and "Professor of Worms and Insects in Paris," is often credited with offering the first scientific theory of evolution. He promoted the idea that life is constantly emerging from nonliving matter by a process of spontaneous generation. From a modern perspective,

the idea that life could emerge anew on an almost daily basis is hard to grasp, but spontaneous generation was a widespread view up until it was discredited by the elegant experiments of Louis Pasteur (1822–1895). Lamarck's book *Philosophie Zoologique* (1809) argued that, after emerging from inanimate matter, life-forms have an internal drive toward improvement causing them to evolve upward on the *Scala Naturae*, the "ladder of life."

The ladder of life (also called the Great Chain of Being), a long-standing concept tracing back to ancient Greek philosophy, is based on the idea that living beings are organized in an ascending series. Organisms that were perceived to be simpler, such as plants, sponges, and worms, were placed lower on the ladder, and organisms like lizards, dogs, and humans were perched higher in accordance with the sense that they are more complex and thus "superior." We now understand that any such ranking of living species is invalid, but given human perceptual biases, it probably seemed obvious to the ancient Greeks (and others) that less complex organisms are *lower* in some abstract scale of existence.

Lamarck's view implied that different living forms had independent origins at a different time in the past. The forms that are highest now could be seen simply as those that arose early and have had a long time to progress upward.

At the same time, and somewhat enigmatically, Lamarck seemed to realize that life did not all fit on one ladder. This may be why his diagram of the relationship of different animals (Figure 2.1) implies nonlinearity. It is, however, not strictly a tree. There is no root, and some living groups, such as reptiles, are drawn in the interior of the tree, implying that these organisms are steps on the way to "advanced" animals such as mammals and birds. While the associated text, obscure as it is, does mention "branches" and "twigs", it clearly refers to progress up a scale of existence rather than descent from common ancestry:

> [This figure] may facilitate the understanding of what I have said. It is there shown that in my opinion the animal scale begins by at least two separate branches, and that as it proceeds it appears to terminate in several twigs in certain places.
>
> This series of animals begins with two branches, where the most imperfect animals are found; the first animals therefore of each of these branches derive existence only through direct or spontaneous generation. (*Philosophie Zoologique*, Lamarck 1809, as translated by Elliot, 1914)

From this quote and associated text we discern two important differences between the thinking of Lamarck and Charles Darwin. First, whereas Lamarck assumed that different kinds of organisms traced to separate origins (by spontaneous generation), Darwin concluded that diverse living species all trace back

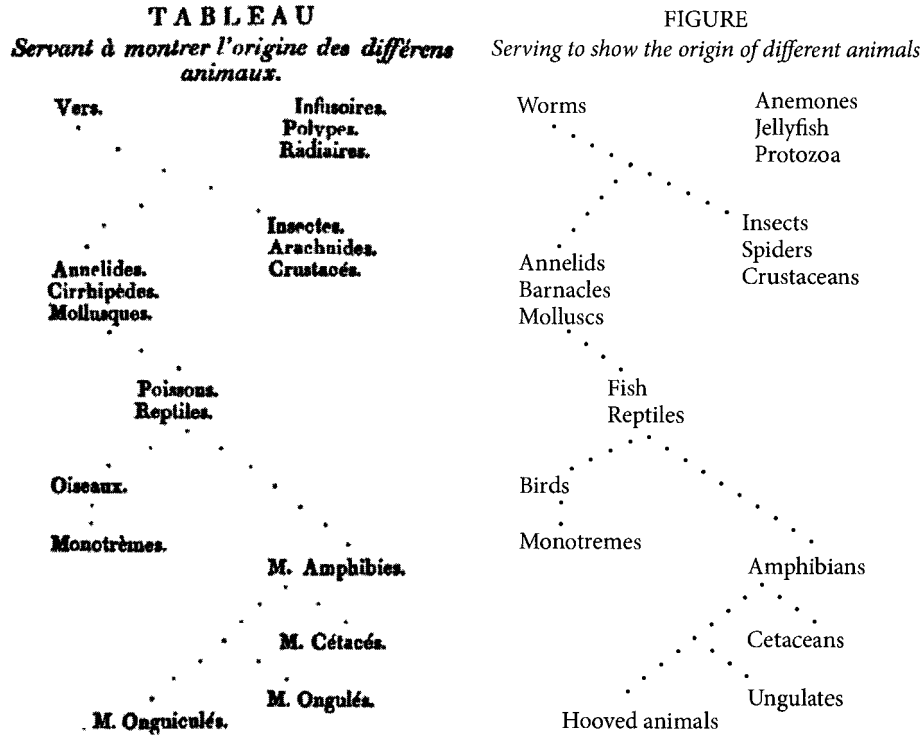


FIGURE 2.1 Lamarck's "tree" of animals. The left panel shows the original French, whereas the right shows it translated into English.

to the same, ultimate common ancestors. Second, whereas Lamarck's thinking seemed built primarily around the ladder metaphor, Darwin developed and made good use of the tree metaphor. How did the transition from separate to common ancestry and from ladders to trees occur?

Around the time of Lamarck, scientific discourse concerning evolution began exploring the concept of shared ancestry. This can be seen, for example, in *Zoonomia*, the lengthy, primarily medical treatise published in 1794 by Charles Darwin's grandfather, Erasmus Darwin (1731–1802).

> . . . would it be too bold to imagine, that in the great length of time, since the earth began to exist, perhaps millions of ages before the commencement of the history of mankind, would it be too bold to imagine, that all warm-blooded animals have arisen from one living filament. . . . (Darwin 1794, Sect IV.8)

In proposing that all warm-blooded animals arose from one ancestral "filament," Erasmus implies some kind of evolutionary branching process that allows one ancestor to yield many, very different descendants.

Not long after the publication of *Zoonomia*, the great English geologist Sir Charles Lyell (1797–1875) provided what may have been the first piece of formal thinking about evolutionary common ancestry. In describing (and discrediting) Lamarck's ideas on evolution, Lyell provided a theory of common ancestry and branching ("ramification").

> We know that individuals which are mere varieties of the same species, would, if their pedigree could be traced back far enough, terminate in a single stock; so according to the train of reasoning before described, the species of a genus, and even the genera of a great family, must have had a common point of departure. What then was the single stem from which so many varieties have ramified? (Lyell 1832, p. 10)

Lyell was pointing out that a valid evolutionary theory would imply that very different living forms, if traced back from offspring to parent to grandparent, and so on, would ultimately converge on a common ancestor. However, the idea that there could be some organism that was a direct ancestor of both a mouse and a human was just too hard to accept. Given the challenge that common ancestry posed to imagination and to religious doctrine, Lyell tentatively rejected common ancestry and, with it, evolution. It is worth noting that while Lyell elegantly describes the principle of common ancestry, he did not, so far as we know, ever draw a tree. It was left to Lyell's protégé, Charles Darwin, to make that critical step.

CHARLES DARWIN'S ROLE IN THE DEVELOPMENT OF TREE THINKING

Charles Darwin (1809–1882), English naturalist, codiscoverer of the mechanism of natural selection, and founder of modern evolutionary biology, was heavily influenced by the work of Lyell, and presumably also by his grandfather Erasmus Darwin. Therefore, it is not surprising that when he pondered the origins of biological diversity, Charles Darwin was drawn to the tree metaphor rather than the ladder of life. Darwin's sketch of 1837 (Figure 2.2) represents a key moment in the development of his evolutionary views. The metaphor he developed is described beautifully in *On the Origin of Species*:

> The affinities of all the beings of the same class have sometimes been represented by a great tree. I believe this simile largely speaks the truth . . .

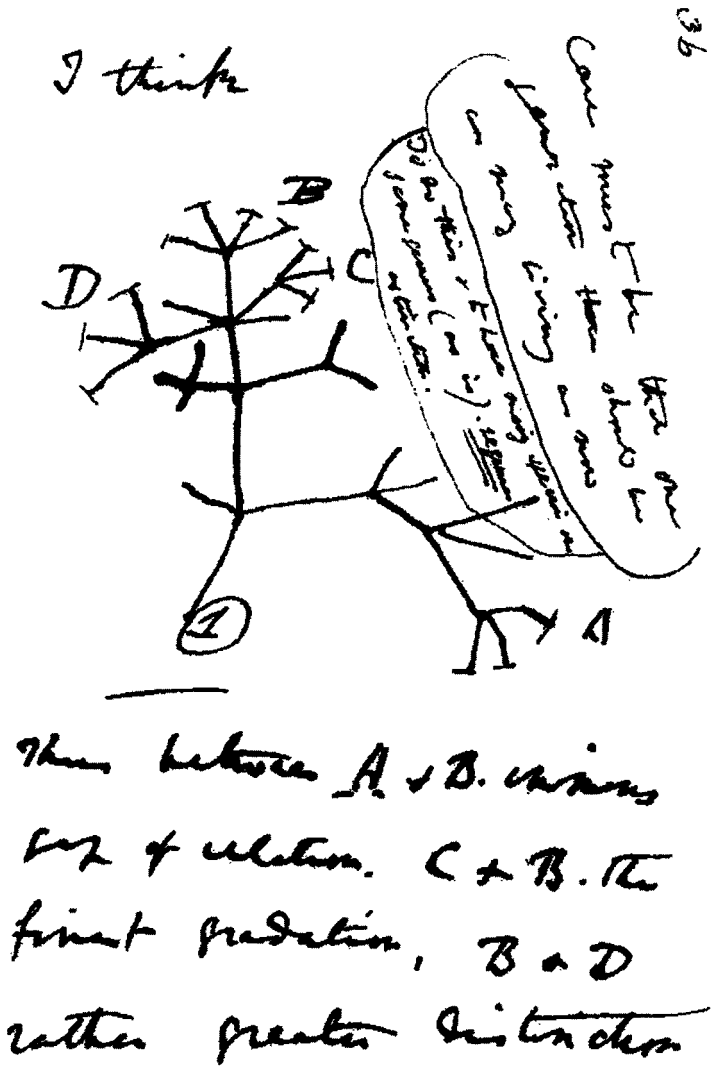


FIGURE 2.2 Charles Darwin's 1837 sketch (Notebook B, entry 36) showing his emerging idea of descent from common ancestry. The text above reads, "I think. Case must be that one generation then should be as many living as now. To do this & to have many species in same genus (as is) *requires* extinction." Text below reads, "Thus between A & B immense gap of relation. C & B the finest gradation, B & D rather greater distinction."

> . . . The green and budding twigs may represent existing species; and those produced during former years may represent the long succession of extinct species. . . .
>
> . . . the great Tree of Life . . . , which fills with its dead and broken branches the crust of the earth, and covers the earth with ever-branching and beautiful ramifications (Darwin 1859, p. 159)

Darwin visualized evolution as having a tree form, with all living species (the tips of branches) connected to one another at branch points (corresponding to common ancestors) that represent organisms that lived in the past.

Darwin, like Lyell, saw that if distinct living species trace back to a common ancestor, then evolution must have happened. If a mouse and a human share a common ancestor, then all of the traits that differentiate these two living species must have accumulated since the time of their common ancestor (Figure 2.3). If common ancestry holds, then evolution must have occurred. Only by arguing that each living kind arose independently in its current form can one deny the fact of evolution. This means that evidence of common ancestry is also evidence of evolution.

On the Origin of Species is often characterized as one long argument for the efficacy of natural selection. Actually, the book places about equal emphasis on the evidence for common ancestry. Almost every chapter includes some statement to the effect that a particular biological fact would be hard to comprehend

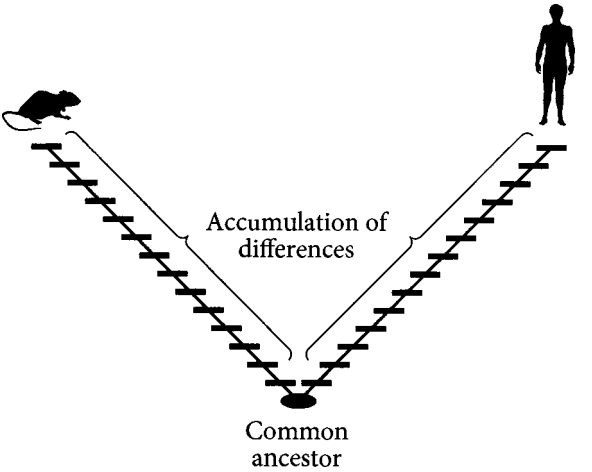


FIGURE 2.3 Common ancestry implies evolution. The bars represent the many traits that have accumulated on the human and mouse lineages since these two species shared common ancestry.

if all living species had been separately created. Indeed, the theory of natural selection is explicitly developed as a follow-up to the basic claim that species have evolved:

> In considering the Origin of Species, it is quite conceivable that a naturalist . . . might come to the conclusion that each species had not been independently created, but had descended, like varieties, from other species. Nevertheless, such a conclusion, even if well founded, would be unsatisfactory, until it could be shown how the innumerable species inhabiting this world have been modified. . . . (Darwin 1859, p. 3)

On the Origin of Species contained a cogent argument for descent from common ancestry, with natural selection offered as an essential part of the argument to explain how species "have been modified." Because of the intimate relationship between common ancestry and tree thinking, it is appropriate to digress briefly to summarize the evidence for evolution and common ancestry that Darwin so thoroughly documented.

EVIDENCE FOR EVOLUTION AND COMMON ANCESTRY

Before we can discuss evidence for common ancestry we need to articulate the competing theory against which it is pitted. The alternative to common ancestry is separate ancestry, which claims that different kinds of organisms arose independently. For our purposes it does not matter whether the cause of the separate origins was by natural or supernatural means, nor whether kinds were created all at once or at different times. The key point is that the theories of separate and common ancestry make different predictions about the kinds of biological facts we should observe. An observation provides evidence for common ancestry if it is likely to have arisen given common ancestry, but unlikely under separate ancestry.

Table 2.1 summarizes the main classes of evidence for common ancestry. Both common ancestry and separate ancestry can account for similarities between organisms when traits are adapted to the same function. However, only common ancestry can readily explain why structures in distantly related organisms can have deep similarities despite fulfilling very different functions. The common bone arrangement in the forelimbs of horses, bats, and humans (Figure 2.4) is found despite the radically different functions of those limbs. This fact is easily explained by common ancestry but enigmatic, to say the least, under separate ancestry. The fact that the same genetic code is used by

TABLE 2.1 Some classes of evidence for evolution and common ancestry

Structural similarities	Traits of different species show deep structural similarities, even when adapted for different lifestyles and even when the traits seem to lack any current utility.
Molecular similarities	Different living species have numerous molecular similarities including both general features, such as use of the same nucleic acid and protein "alphabets," and specific features, such as occurrence of genes with extremely similar composition.
Geographical distributions	Species tend to live in the same part of the world as their close relatives and often do not occur in climatically similar but distant regions. Fossils tend to be found in the same part of the world as their living relatives.
Fossils	Older fossils differ more from living species than do younger fossils. Some transitional fossils appear intermediate in form between living species.
The treelike distribution of traits among species	Classifications are structured with nonoverlapping groups consistent with descent from common ancestry along the branches of a tree. Traits show more treelike structure than would be expected by chance under separate ancestry. Different genes from the same organisms tend to suggest the same tree.

the bacterium *Escherichia coli* and humans can hardly be explained by such a code being optimal for building such different organisms. Likewise, there are numerous examples of vestigial traits: traits that seem to have lost an ancestral function, but are still present. For example, the presence of a pelvic bone in whales or chloroplasts (plastids) in nonphotosynthetic plants is readily understood under the hypothesis that they share common ancestry with walking mammals and photosynthetic plants, respectively.

The evidence from the fossil record is sometimes misunderstood. The mere fact that fossil forms differ from living species or from fossils in other strata does not, in itself, provide evidence for common ancestry. It is, *a priori*, possible that these different forms were independently created. Unless you can establish that one fossil individual is an ancestor of multiple, different-looking

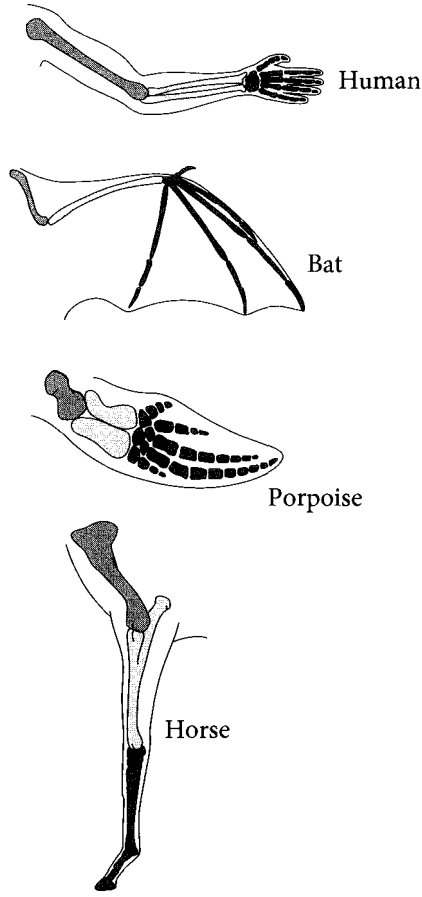


FIGURE 2.4 Homologous features. The forelimbs of these four species differ in relation to their lifestyle but share the same set of bones.

descendants, which is very hard to do, you can explain fossils as independently created, but now extinct forms.

Fossils become evidence for evolution when we find patterns that are predictable from the principle of common ancestry, but not predictable from separate ancestry. Two patterns, both noted by Darwin, are worth mentioning. First, if fossils are part of the same great tree of life as living forms, then those forms that went extinct a long time ago should be more different from living species than are those that went extinct only recently.

> On the principle of the continued tendency to divergence of character . . . we can understand the rule that the most ancient fossils differ most from existing forms. (Darwin 1859, p. 331)

Figure 2.5 represents evolutionary lineages overlaid upon some fossil-bearing rocks. Lower layers of rock contain older fossils than higher layers. Darwin's argument is that if fossils *c*, *d*, and *f* all share more recent common ancestry with living species B than with any other living species, we expect a graded degree of similarity, for example, with *f* being more similar to *d* than to *c*. Such a pattern in the temporal distribution of fossils is expected given common ancestry and, importantly, is observed in many cases. Under separate ancestry it is unclear why such a pattern would hold.

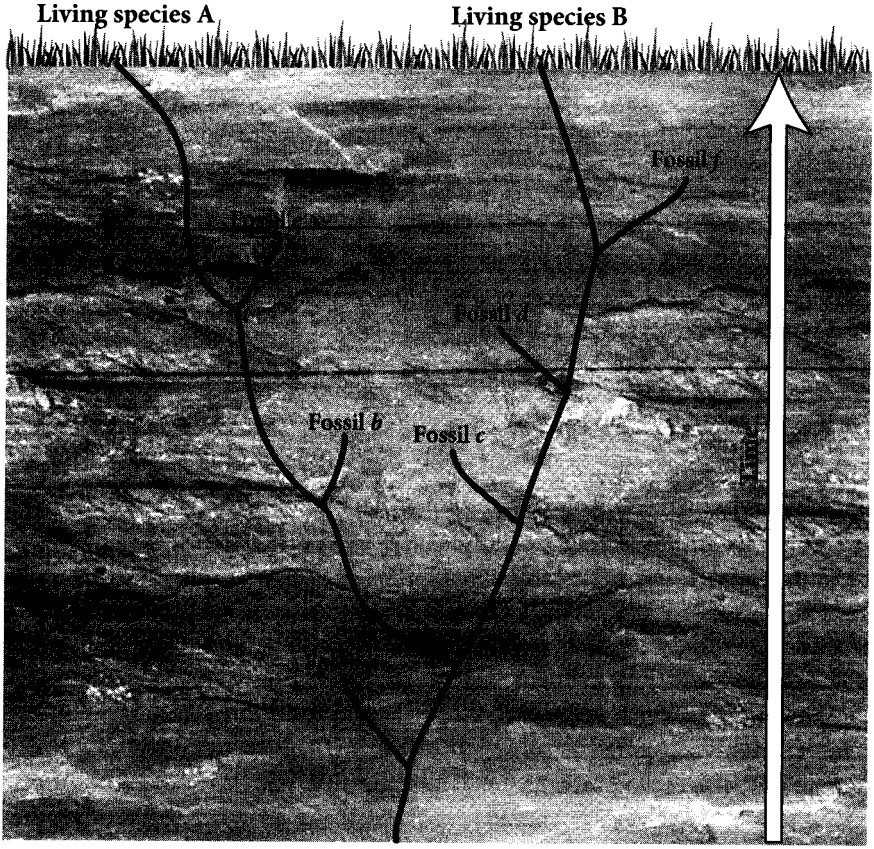


FIGURE 2.5 Relationships of living species to fossils. Changes accumulate over time, making fossils *e* and *f* more distinct than fossils *b* and *c*.

Likewise, common ancestry predicts that, as one goes backwards through time, the difference between two fossil groups will tend to diminish.

> [T]hose groups which have, within known geological periods, undergone much modification, should in the older formations make some slight approach to each other; so that the older members should differ less from each other in some of their characters than do the existing members of the same groups; and this by the concurrent evidence of our best palaeontologists is frequently the case. (Darwin 1859, p. 333)

Referring to Figure 2.5, because changes will have accumulated along the lineages leading to living species A and B, we have reason to predict that the ancient representatives of the two lineages (e.g., *b* and *c*) will differ less from each other than will later occurring representatives (*e* and *f*, or A and B). This pattern, familiar to paleontologists, is predictable from the principle of common ancestry but inexplicable under separate ancestry.

An extension of this way of thinking about the fossil record is the prediction that we will find transitional fossils: fossils that manifest some but not all of the distinctive traits of a living group. If descent from common ancestry holds, and if evolution is a gradual process, then for every living group that has numerous unique traits there must have once existed organisms that possessed a subset of these traits. Transitional fossils are not predicted under separate ancestry, so the many transitional fossils discovered by paleontologists provide compelling evidence in favor of common ancestry and evolution.

There is also geographic evidence for common ancestry, which applies both to fossil and living forms. Darwin was struck by the similarities between fossil and living species from the same part of the world. Early in the voyage of the *HMS Beagle* he found some gigantic glyptodont fossils, which only occur in South America, and noted that they show striking similarities to armadillos, which also hail from this continent. This is expected if glyptodonts and armadillos share a recent common ancestor that lived in South America and if the rate of migration around the world is low. This would not be predictable under separate ancestry—structurally similar fossil and living species could as easily be found on opposite sides of the globe.

> On the theory of descent with modification. . . . succession of the same types within the same areas, is at once explained; for the inhabitants of each quarter of the world will obviously tend to leave in that quarter, during the next succeeding period of time, closely allied though in some degree modified descendants. (Darwin 1859, p. 340)

The same logic applies to living groups. Have you ever considered why polar bears don't live in Antarctica and penguins don't live in the Arctic? Given the climatic similarity of these regions, why should one find such a pattern? Like-

wise, why are there no hummingbirds in African or Asian rainforests, no cacti native to African deserts, no kangaroos outside Australasia, and so on?

Common ancestry explains these and many other distribution patterns seen in ***biogeography***. Except in a few highly mobile groups or in groups inhabiting landmasses that have moved greatly over the eons, offspring will "inherit" the geography of their parents. Therefore, we expect closely related organisms to live close to one another. As a result, whole branches of the tree of life will tend to be geographically constrained. If bears originated in the northern hemisphere (as seems likely based on phylogenetic analyses) we can understand why they were able to invade the Arctic, but have not dispersed through the tropics to Antarctica.

The final class of evidence for evolution relates most directly to phylogenetic trees and tree thinking. For this reason, we believe that it is worth summarizing in a bit more detail.

When we look at the distribution of traits among species or at classification systems that are based on these traits, we see clear hallmarks of an underlying treelike form. Biological taxonomic systems, comprising species within genera, genera within families, and so on, have been used since well before Darwin as a way to organize the grand patterns of diversity on the planet. Taxonomies have a nested hierarchical structure: two groups may be nested one within the other but they are never partially overlapping (Figure 2.6). Darwin inferred that this hierarchical structure was not merely a convention but resulted from a true tree

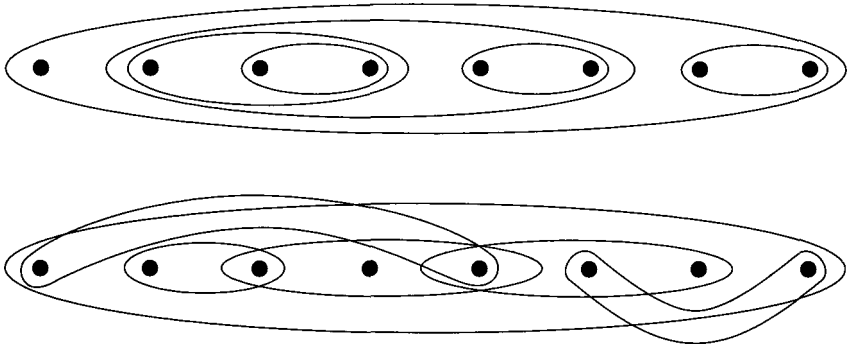


FIGURE 2.6 Comparison of hierarchically nested and overlapping classification systems. The top row shows a nested classification of eight species with the species represented by dots and the taxa they are assigned to represented by ovals. The bottom row shows an overlapping classification with taxa represented by ovals or arch-shapes.

of evolutionary relationships that shaped the traits upon which classifications are made:

> From the first dawn of life, all organic beings are found to resemble each other in descending degrees, so that they can be classed in groups under groups. This classification is evidently not arbitrary like the grouping of the stars in constellations. (Darwin 1859, p. 411)

Darwin understood that such a nested hierarchical structure would arise if there is treelike ancestry and if taxonomic groups are composed of organisms linked by evolutionary kinship. If you classified the leaves on a botanical tree on the basis of the branches to which they were attached, you would obtain a strictly hierarchical classification of leaves. Likewise, classifications that group organisms based on traits that have evolved along an evolutionary tree are hierarchical (see Chapter 5). In contrast, it is all but impossible to see why separate origins would cause traits to support a nested hierarchy of groups within groups:

> On the view that each species has been independently created, I can see no explanation of this great fact in the classification of all organic beings. (Darwin 1859, p. 129)

Although not something that Darwin knew about, statistical analyses of the traits of living species show that the distribution of traits has the unmistakable hallmark of descent from common ancestry along the branches of a tree. The ***consistency index***, or CI (see Chapter 4), measures the extent to which traits conform to a treelike form. We can take a data set and find the tree on which those data have the highest CI. Then we can generate hundreds of data sets that are equivalent to the original data but resemble what we would expect if the species had arisen by separate ancestry (Chapter 9). For each of these simulations we can find the tree that maximizes the CI, and then we can compare this CI to that obtained with the real trait data. Table 2.2 shows some examples collected by James Archie in 1989. He looked at all available published studies and found that every one showed more hierarchical, treelike structure (a higher CI) than would be expected under separate ancestry. These examples illustrate the convincing treelike structure of trait variation regardless of the group of organisms or type of data collected.

The treelike structure found when looking at the physical traits of organisms (Table 2.2) is found even more strongly when analyzing molecular data sets. When we sequence the same gene from multiple species and subject it to this statistical test, we generally find huge differences between the observed CI and the CI predicted under separate ancestry.

In 1982 David Penny and collaborators conducted a statistical test of another prediction of descent from common ancestry. They showed that different genes

TABLE 2.2 An analysis of published morphological data showing that the amount of tree structure, as measured by the consistency index (CI), is significantly greater than expected under separate ancestry (based on Archie, 1989)

Organisms studied	Observed CI	CI expected under separate ancestry
Anacyclus (plants)	0.81	0.62
Angiosperms (plants)	0.21	0.16
Peromyscus (mammals)	0.76	0.57
Chloris (plants)	0.51	0.42
Cnemidophorus (reptiles)	0.32	0.23
Corrella (birds)	0.95	0.63
Dasyuridae (mammals)	0.23	0.17
Dipodomys (mammals)	0.58	0.48
Drosophila 1 (insects)	0.30	0.17
Drosophila 2 (insects)	0.83	0.38
Equus (mammals)	0.81	0.60
Gerygone (birds)	0.42	0.26
Heliconia (plants)	0.70	0.26
Leptodactylidae (amphibians)	0.22	0.16
Leptopodomorpha (insects)	0.85	0.56
Menidia (fish)	0.79	0.40
Microteiidae (reptiles)	0.44	0.38
Myobatrachidae (amphibians)	0.35	0.22
Opheodrys (reptiles)	0.68	0.54
Orthoptera (insects)	0.51	0.39
Percina 1 (fish)	0.40	0.27
Percina 2 (fish)	0.90	0.32
Podarcis (reptiles)	0.28	0.19
Pomacentridae (fish)	0.67	0.37
Pygopodidae (reptiles)	0.59	0.23
Salamandridae (amphibians)	0.60	0.44
Tamias (mammals)	0.59	0.48
Uma (reptiles)	0.86	0.72

sampled from the same species yield the same tree, or if not exactly the same tree (perhaps because of errors in tree reconstruction) then trees that are much more similar to one another than would be expected given separate ancestry. Penny et al. suggested that this pattern is so compelling that the hypothesis of separate ancestry should be emphatically rejected as being incompatible with the data. While science rarely deals with certainty, it is fair to say that evolution from common ancestry is now supported beyond any reasonable doubt.

THE EMERGENCE OF PHYLOGENETIC SYSTEMATICS

Darwin's work, together with that of other nineteenth-century naturalists, quickly convinced the scientific community that living species are all connected to one another by lines of common descent. While it took approximately eighty years for natural selection to be accepted as the main driver of adaptive evolution, the fact of common ancestry was acknowledged almost immediately. Interestingly, while naturalists accepted that living species could be visualized as the tips of the great tree of life, the full implications of this branching structure did not emerge until more than a century after the publication of *On the Origin of Species*. Two factors likely played a role in this delay.

First, as you will see in the upcoming chapters, tree thinking is conceptually challenging. It is easy enough to accept trees at some level while still retaining the vestiges of ladder thinking. A good indication of the prevalence of this hybrid mentality was the common practice, well into the 1970s, of organizing living species into "phyletic series." In such series, some living groups are implicitly identified as ancestors of others. While the resulting diagrams, such as Bessey's (1915) summary of flowering plant relationships (Figure 2.7), look like trees, some living groups appear to be descended from other living groups. This is, of course, impossible if the tree is taken to show lines of actual descent. However, the diagram is intended to communicate information on "advancement"—a loosely defined concept that derived ultimately from the ladder of life philosophy. Thus, Figure 2.7 suggests that the "primitive" Rosales gave rise to Celastrales, which gave rise to Umbellales (e.g., carrots), then Rubiales (e.g., coffee), and finally the most "advanced" order, Asterales (e.g., daisies). There is an implication of a ladder of advancement along one or a few major axes.

A second reason for the failure of tree thinking to emerge was a lack of practical tools for reconstructing trees. There was a pulse of efforts to infer phylogenies using embryological data in Germany in the late nineteenth and early

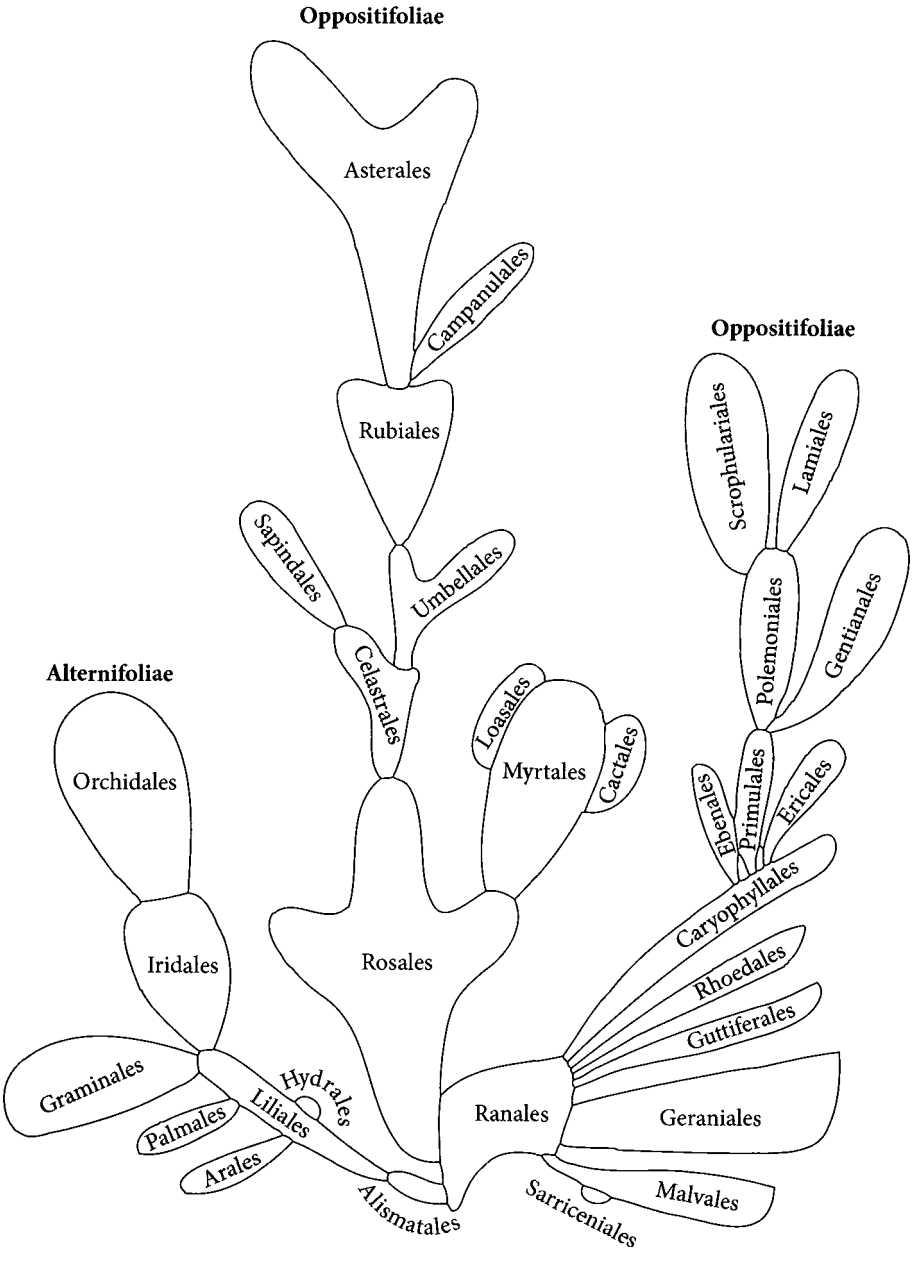


FIGURE 2.7 Bessey's phyletic series of flowering plants. Living groups (plant orders) are arranged in a series of ancestor-descendant relationships. This visualization conflicts with the nature of evolution in which living groups do not give rise to other groups but are related by common ancestors.

twentieth centuries. Ernst Haeckel (1834–1919), in particular, generated a series of beautifully illustrated phylogenies such as that reproduced in Figure 2.8 (note the implicit ladder of advancement with humans at the pinnacle of the tree). The hope was that embryological data could be filtered through a set of formal laws to yield unambiguous insights into phylogenetic history. In practice, while some valid conclusions were reached, embryological data were difficult to collect, and obtaining phylogenies from such data was far from straightforward.

Although the idea that embryology precisely replicates phylogenetic history eroded through the first half of the twentieth century, the embryological tradition is likely to have provided the foundation for the German pioneers of phylogenetic theory: the entomologist Willi Hennig (1913–1976) and the botanist Walter Zimmerman (1892–1980). From the 1930s to the 1960s, they developed a way of thinking about evolutionary trees and their relationship to taxonomic systems, which they called ***phylogenetic systematics***. Three aspects of their views were particularly important and have come to define the modern field of systematics.

First, phylogenetic systematists argued for the objective reality of evolutionary trees and the common ancestors that they imply. The aim of systematic research, they suggested, is to elucidate the true tree using observations of the traits of organisms. However, we should never expect to achieve perfect knowledge of the tree of life. Phylogenetic systematics is a science, and like any other science, we can make well-supported inferences about the truth but we can never mathematically prove that those inferences are true.

Second, phylogenetic systematists proposed that the degree of relatedness among organisms should be understood in terms of recency of common ancestry. While we might be tempted to take account of how similar two organisms are when we assess their relationship, phylogenetic systematists argued that we should consider *just* their evolutionary kinship—how recently they last shared a common ancestor. Because unrelated lineages can converge on the same traits and because rates of evolution differ between lineages, similarity and relatedness may be at odds.

Third, and most controversially, phylogenetic systematists argued that phylogenetic relatedness should be the sole basis of biological classification. The standard practice in taxonomy at the time was to classify species using a combination of inferred relationships and degree of similarity. The phylogenetic approach, in contrast, holds that only evolutionary relationships count. If a

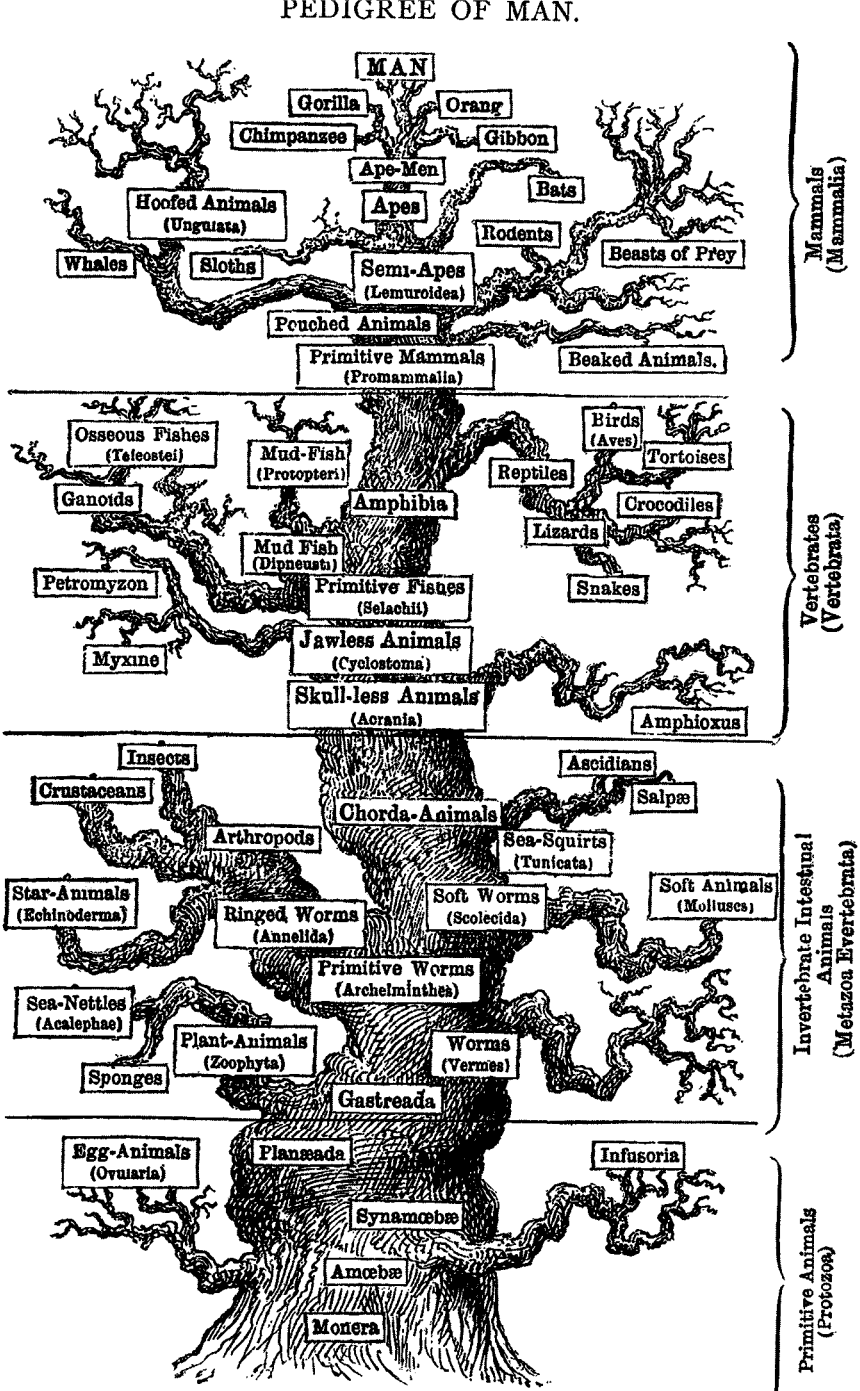


FIGURE 2.8 Haeckel's tree of life. Some living groups are shown along branches, while others appear at the tips.

crocodile is inferred to have had a common ancestor with a bird more recently than with a lizard, then the crocodile should be classified closer to a bird than to a lizard. Regardless of the fact that a crocodile and a lizard share many features that are missing in birds, phylogenetic systematists would not recognize a taxon that included lizards and crocodiles unless it also included birds.

The publication in English of Hennig's book, *Phylogenetic Systematics* (1966), served to introduce the non-German-speaking world to this phylogenetic perspective. It is difficult to overstate the intensity of the debate that erupted. In particular, the argument that classification should be based solely on degree of relatedness was vehemently opposed. The vehemence of the response is not surprising, since the phylogenetic approach called into question the reality of entities that were almost universally perceived as being real. Most famously, phylogenetic systematists questioned the validity of the vertebrate class Reptilia (an example explored further in Chapter 5). Reptilia (reptiles), then understood to comprise turtles, tuatara, snakes, lizards, and crocodiles, seemed to be as real as Mammalia (mammals) or Aves (birds). Phylogenetic systematists, or as they were sometimes called, "cladists," accepted the reality of Mammalia and Aves, because each group is composed of organisms that show a closer evolutionary relationship to other members of the group than to any organisms outside the group. However, they rejected the reality of the traditionally recognized Reptilia because some reptilians (e.g., crocodiles) are more closely related to nonreptilians (e.g., birds) than they are to some other reptilians (e.g., lizards). By denying the validity of a familiar group such as Reptilia, phylogenetic systematics had very much thrown down the gauntlet.

The traditional systematists, or as they renamed themselves "evolutionary systematists," hit back by arguing that classification should take account of both ancestry *and* similarity. As an example, they argued that Reptilia could be considered a "natural group" because it is composed of organisms that share common ancestry, look similar, and have similar ecologies. The phylogenetic systematists responded by pointing out that there is no one meaningful way to quantify similarity. And even if there were, the decision as to how much similarity is needed to outweigh ancestry would be subjective. Their argument was weakened by the fact that, at the time, it was easier to measure similarity than to confidently determine evolutionary relationships. The choice was between classifying organisms based on similarity, a subjective but tangible and directly observable criterion, versus phylogenetic kinship, an objective but difficult-to-measure criterion.

This impasse was resolved in the 1980s and 1990s with great improvements in methods for inferring phylogenies. These advances were made possible by the development of new statistical methods, the application of improved computers, and the easy availability of DNA sequence data. Thanks to these developments, systematists made some startling discoveries about evolutionary history. With this quantum improvement in our knowledge of evolutionary relationships, the phylogenetic systematic approach to classification gradually prevailed. As explored in Chapter 5, the dominant view today is that classification should be based on common ancestry alone.

VESTIGES OF LADDER THINKING IN THE POPULAR UNDERSTANDING OF EVOLUTION

Improvements in phylogeny reconstruction at the end of the twentieth century not only changed the theory and practice of taxonomy but also catapulted trees into a place of prominence within evolutionary biology. At first, evolutionary biologists used phylogenetic trees as tools, but made only modest changes to the way that they conceptualized evolution or communicated evolutionary facts to the public. Still, after twenty years of tree-rich evolutionary biology, ladder thinking (or what Robert O'Hara (1997) called "development thinking") continues to permeate the language and visual representation of evolution in the literature and in the media.

To illustrate this point, one can start with any number of images that depict some living species as being evolutionary ancestors of another living species. There are many cartoons about human evolution that depict us descending from chimpanzees (e.g., Figure 2.9). As amusing as these cartoons may be, we know that humans did not evolve from modern chimpanzees, any more than chimpanzees evolved from humans. Rather, the last common ancestors of humans and chimpanzees were neither humans nor chimpanzees.

Another indication of ladder thinking is the language that scientists or journalists tend to use when telling the story of an evolutionary transition. For example, consider the invasion of land by vertebrates. This is often told in the form of a linear narrative: fish gave rise to amphibians. The paleontological challenge is then cast as the search for an intermediate between a fish and an amphibian. If one focuses on the traits implied in this story, it is accurate: fish traits such as fins are the structural progenitors of amphibian traits, such as walking limbs. It is also valid to say that an ancestor with a body form and

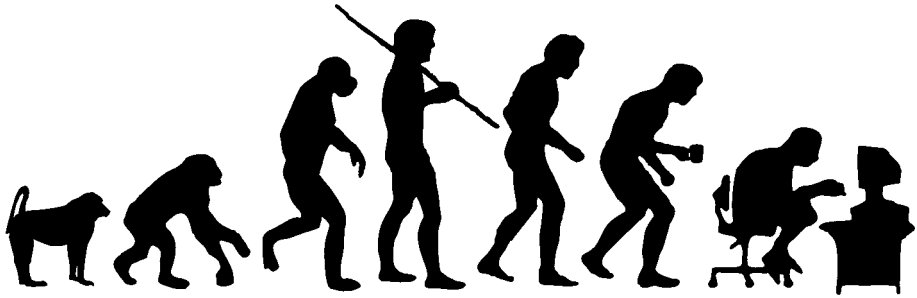


FIGURE 2.9 Ladder-thinking cartoon. This image perpetuates the incorrect "ladder-thinking" notion that humans evolved from living species such as baboons and chimpanzees.

ecology like a modern fish gave rise to a descendant that could walk on land like an amphibian. However, to most people the story implies a progression among a set of living organisms: ray-finned fish (e.g., goldfish, salmon, and tuna) giving rise to lobe-finned fish (lungfish and coelacanths) giving rise to amphibians (e.g., salamanders). While nobody telling the story of tetrapod evolution really thinks that a living goldfish is an ancestor of a living amphibian, there is still an incorrect implication that the living "fish," as a group, are ancestral to amphibians and, as a result, that fish have evolved less than amphibians. However, both living fish and living amphibians have evolved exactly the same amount of time from their common ancestor and both have acquired new traits during that time.

One manifestation of ladder thinking is calling some living organisms "ancient," "primitive," or "lower." The intention may be to say that the group in question is little changed in external appearance from certain ancestors, but the statement mistakenly implies that one taxon has only ancestral traits. This is never true. Just because a living lamprey is little changed in overall form from the early ancestors of all living vertebrates does not make living lampreys the actual ancestors. And while they might bear the ancestral condition for many visible traits, there is every reason to assume that, for some traits, lampreys have an evolutionarily derived condition, whereas other vertebrates—for example, humans—have the ancestral condition. For instance, lampreys are unique among living vertebrates in having a single, medial nostril.

One reaction to the persistence of ladder thinking in the terminology and imagery of evolution is to let it slide. So long as people understand the branching, treelike structure of evolutionary history, what harm is done by language

that implies the existence of a ladder of life? The problem is that most people do not understand the treelike nature of evolution, and constant exposure to ladder-laden language prevents them from learning it. Consider, for example, when TV interviewer Larry King on August 23, 2005, asked his guest Barbara Forrest, " If evolution is true, why are there still monkeys?" Odd as this question may sound to an evolutionary biologist, Larry King could not be faulted for asking it if he has frequently heard biologists say, "Humans evolved from apes, which evolved from monkeys." If, instead, human evolution were consistently described using more accurate language, "Hominids are a subgroup of apes, and apes are a subgroup of monkeys," then Larry King and others like him might be less confused over the basic concepts of evolution.

As we look to the future, we hope that trees rather than ladders will come to be the main evolutionary metaphor, because such a change will inevitably signal a deeper and more widespread understanding of the fundamental tenets of evolution. Responsibility for achieving this transition lies with three constituencies. Biologists need to become more careful in the language that they use to describe evolutionary transitions to avoid implicitly endorsing ladderlike, progressive interpretations. Science educators and journalists need to confront tree thinking more deliberately and to emphasize tree-thinking skills as critical learning goals for biological literacy. The educated general public should make the effort to learn the underlying concepts of tree thinking and to use them to achieve a more nuanced and accurate understanding of evolution and biological diversity. We hope that the text, figures, and problem sets in this book will advance this objective.

FURTHER READING

Early evolutionary and systematic theory: Panchen 1992; Bowler 2003; Browne 2006

Evidence for evolution: Mayr 2001; Coyne 2009; Dawkins 2009; Zimmer 2009

Tests for the significance of tree structure: Penny et al. 1982; Archie 1989; Steel and Penny 2010; Theobald 2010

Phylogenetic systematics and the cladistic revolution: Hennig 1966; Mayr 1982; Hull 1988

Tree thinking versus ladder thinking: O'Hara 1992, 1997; Crisp and Cook 2005; Omland et al. 2008

CHAPTER 2 QUIZ

1. What is meant by "common ancestry?"
   a. Common species are ancestors of rare species
   b. All organisms are ancestors
   c. Very different living species descend from the same ancestor
   d. Simple living species are ancestral to more complex species
   e. There is a ladder of life with primitive and more advanced species arrayed on different "rungs"

2. Why is common ancestry important in evolutionary theory?
   a. If common ancestry is true, then species are all created equal
   b. Only if common ancestry is true could evolution be viewed as progressive
   c. If common ancestry is true, then the idea of special creation gains some support
   d. If common ancestry is true, then evolution (change over time) must have happened
   e. Common ancestry shows that natural selection is the only mechanism of evolution

3. Which of the following facts provides the clearest evidence for common rather than separate ancestry?
   a. The coelacanth fish looks almost indistinguishable from fossils in 200-million-year-old rocks
   b. Most primates have tails, which seem to be important for their survival
   c. Diverse cactus species are found in the American deserts, but none occur in African or Asian deserts
   d. Some orchid flowers are very well suited to pollination by particular kinds of insects
   e. Whale flippers and dolphin flippers have a similar bone arrangement and are used for a similar function

4. What are vestigial organs?
   a. Traits whose similarity is due to common ancestry (like a horse foreleg and a human hand)
   b. Traits that are functional but poorly "designed" (like the vertebrate blind spot)
   c. Traits that look similar but evolved independently (like a bird wing and a bat wing)
   d. Traits that evolved by a mechanism other than natural selection (like human altruism)
   e. Traits that have become reduced and nonfunctional (like the thigh bones of pythons) but are functional in related species.

5. Which of the following statements is most clearly aligned with tree thinking rather than ladder thinking?
   a. Humans are a kind of ape
   b. Jellyfish are more primitive than goldfish
   c. Birds evolved from primitive reptiles, like dinosaurs
   d. Lower animals lack backbones
   e. Dolphins are the most advanced animals

6. Which of the following fossil discoveries would provide the most compelling evidence that primates and rodents descended from a common ancestor?
   a. In strata from 70 Ma (million years ago) you find a fossil almost identical to living primates and another that is almost identical to living rodents
   b. A fossil from 40 Ma has general primate features but also has unique traits that do not occur in any living primates
   c. Fossil primates are found in Northern Europe where no primates currently live, but where rodents do currently occur
   d. A series of fossils is found between 70 Ma and 80 Ma that have traits shared by primates and rodents, with later ones tending to have either specific rodent or specific primate traits
   e. A fossil from 10 Ma has quills identical to porcupines (a kind of rodent) and a prehensile tail that is identical to New World monkeys

7. Which of the following is predicted under common ancestry but not under separate ancestry?
   a. Gene sequences will tend to vary among species so that no two species have exactly the same sequence of bases in their genome
   b. The gene sequences from two very different species can be lined up in such a way that some bases match
   c. The tree constructed for a group of species will be similar or identical, regardless of which gene sequences were used to build the tree
   d. Gene sequences encode proteins that will be functional only in those organisms in which they are found
   e. Two species of insects that can tolerate the same pesticide will use different biochemical mechanisms to detoxify the pesticide

8. From the 1960s through the 1980s, the views of phylogenetic systematists such as Hennig and Zimmerman were opposed by evolutionary systematists. Which of the following was seen as the most controversial claim of phylogenetic systematics?
   a. The idea that some species are more advanced than others
   b. The idea that organisms as different as insects and plants descend from common ancestry

c. The idea that evolutionary relatedness should be taken into consideration in developing a biological taxonomy
d. The idea that phylogenetic trees can be reconstructed from the traits of living organisms
e. The idea that overall similarity should not be used as a basis for biological classification

9. What factors explain the increasing importance of phylogenetic trees in evolutionary biology toward the end of the twentieth century?
   a. Improved statistical methods for phylogenetic analysis
   b. Enhanced availability of molecular data that could be used for phylogenetic inference
   c. Advances in computer power that made it easier to infer phylogenetic relationships
   d. Increasing appreciation for the importance of communicating the treelike nature of evolutionary history
   e. All of the above

10. Refer to Figure 2.5. Which of the following pairs would you expect to have the greatest degree of difference from each other?
    a. Living species A and living species B
    b. Living species A and fossil *c*
    c. Living species B and fossil *c*
    d. Fossil *b* and fossil *d*
    e. Fossil *e* and fossil *f*

11. What are "transitional fossils" and how do they provide evidence of evolution?

12. What are the essential differences between tree thinking and ladder thinking?

13. In what ways did the evolutionary views of Lamarck and Charles Darwin differ?

14. How would a phylogenetic systematist make sense of the conclusion that crocodiles are more closely related to birds than to lizards, given that lizards and crocodiles look so similar to each other and so different from birds?

15. Statements based on an inaccurate perception of the process of evolution abound in the media and even in older textbooks. For example, one textbook states, "Lower plants [e.g., mosses, liverworts] evolved into higher plants [e.g., flowering plants]." How does this statement misrepresent evolution?

CHAPTER THREE

What a Phylogenetic Tree Represents

Before you can use trees to organize knowledge of biodiversity and refine your understanding of evolution, you need to know what a tree diagram represents and to become comfortable with some of the conventions used to communicate phylogenetic information. As noted by Robert O'Hara in his seminal exploration of the concept of tree thinking, "just as beginning students in geography need to be taught how to read maps, so beginning students in biology should be taught how to read trees and to understand what trees communicate" (O'Hara 1997). In this chapter, we begin by clarifying how reproduction of individual organisms within populations is connected to ancestry and descent at the level of the tree of all life. We then describe how to read tree diagrams in order to extract their essential phylogenetic content.

THE CONTINUITY OF REPRODUCTION FROM THE POPULATION TO THE TREE OF LIFE

A phylogenetic tree depicts the evolutionary ancestry of a set of tips. The tips are typically living species or groups of species, but can also be fossil organisms, individual organisms, genes, or populations. There is only one basic kind of ancestry in biology: that which links parents, or more generally ancestors, and their children, or descendants. A pedigree or family tree is a depiction of the ancestor-descendant relationships within a population. A phylogenetic tree shows the same thing, but at a larger scale and in less microscopic detail. To understand what a phylogenetic tree depicts, therefore, we need to conceptually bridge the reproduction of organisms within populations and the branching of evolutionary lineages to create a phylogenetic tree.

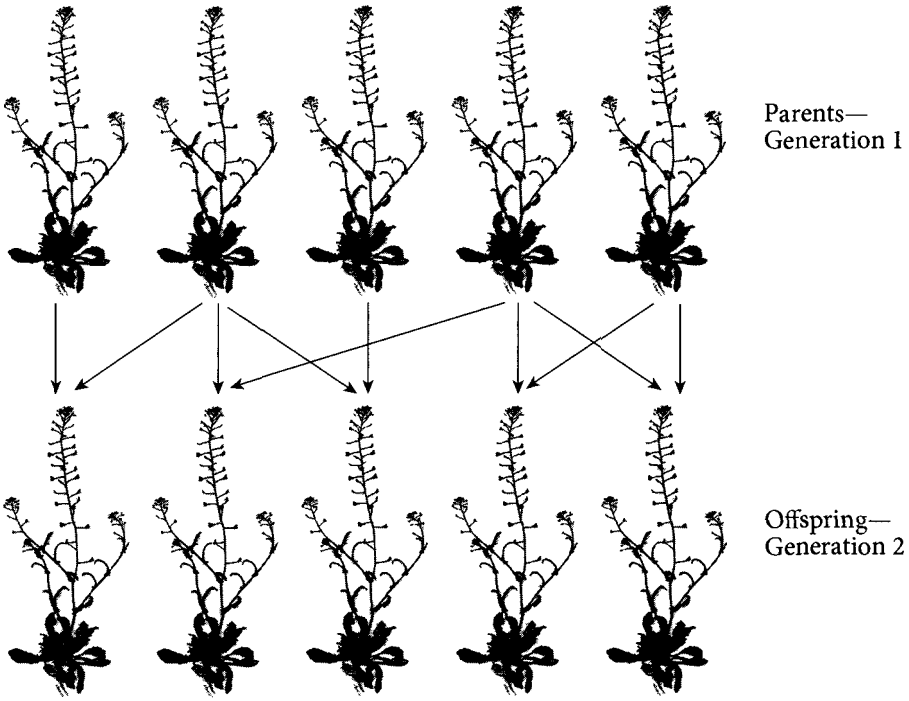


FIGURE 3.1 Two generations of shepherd's purse (*Capsella bursa-pastoris*) plants.

Start by imagining one generation of plants of a particular species; for example, the flowering plant shepherd's purse, growing side by side in a meadow and producing offspring by exchanging pollen. If we focus on five individual plants in the parental generation and the offspring generation, the pedigree could look like that shown in Figure 3.1. Here we have assumed that each individual has two different parents although self-pollination can occur.

If we now expand our image to encompass all plants in a population and several generations, it might look something like Figure 3.2. Notice that each individual has two parents, but gives rise to a variable number of offspring in the next generation.

Imagine taking such a pedigree and getting rid of the organisms so that only the descent relationships were retained, as shown in Figure 3.3. These parent-offspring connections can be thought of as the glue that holds the population

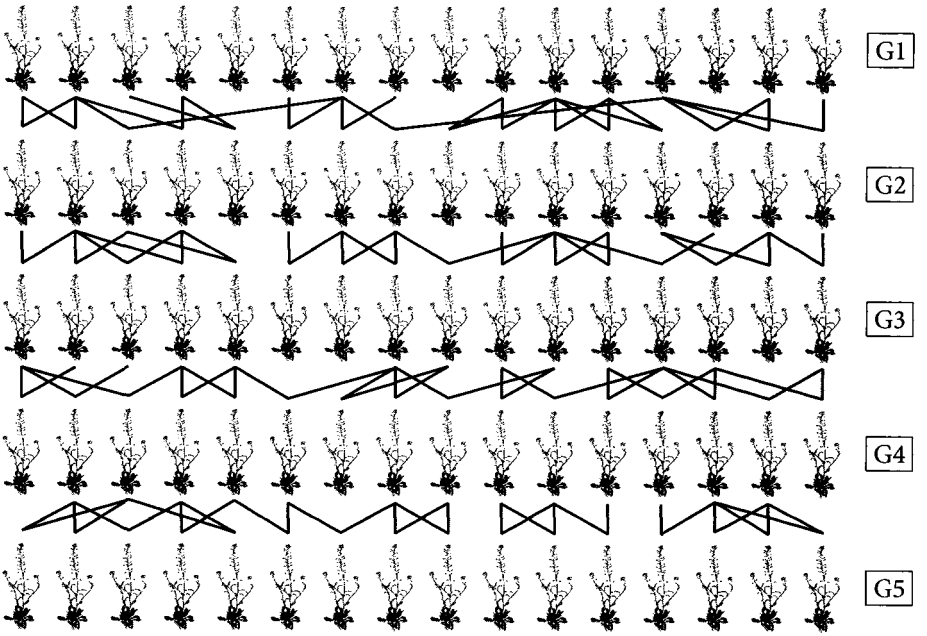


FIGURE 3.2 Multiple generations in the shepherd's purse population.

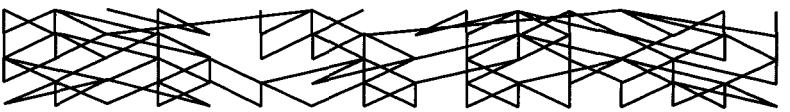


FIGURE 3.3 A representation of the generations in Figure 3.2 showing only lines of descent.

together. When we get to evolutionary timescales, which typically entail hundreds of thousands or millions of generations, individual organisms are too transient to be of concern, except as the vehicles through which the lines of descent pass. The lines of descent are what we most care about.

Instead of visualizing one small part of a single field of shepherd's purse over five years, expand your field of view to include many more individuals and generations. For example, Figure 3.4 is derived from a similar diagram as the preceding but now includes about 250 individuals and 80 generations. If

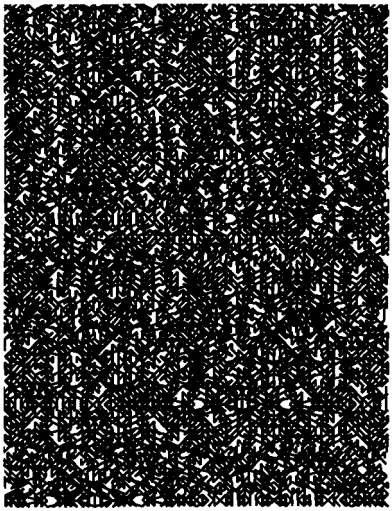


FIGURE 3.4 Many individuals and many generations of the shepherd's purse population.

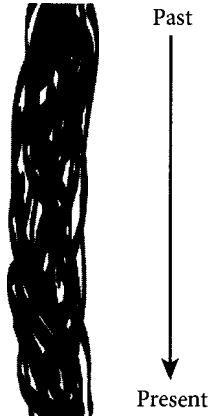


FIGURE 3.5 A lineage containing multiple populations connected by gene flow.

you zoomed out further and tried to represent a typical population of several thousand individuals that persists for hundreds or thousands of generations, all you would be able to see would be a fuzzy line. This line represents the genetic continuity through parent-offspring descent of a single well-demarcated population.

Individual populations may be fairly isolated for some period of time. However, on evolutionary timescales, seeds and pollen will occasionally carry genes between the distinct populations that make up a typical species. This gene flow between populations has the effect of "braiding" population lineages together. The graphic in Figure 3.5 might help you to visualize this braiding. Zooming out still further, this would probably look, again, like a fuzzy line. This is what is usually understood by an "evolutionary lineage."

During evolution, evolutionary lineages may diverge or "split." A technical term for such events is ***cladogenesis***, which refers to the origin (genesis) of new branches (*clados* is Greek for branch). Because lineages are sometimes equated with species, splits of this kind are sometimes called speciation events. However, considering the great controversy as to the exact meaning of "species" (see

Chapter 6), we avoid the term speciation in the following, instead using the less loaded term ***lineage splitting***.

Lineage splitting occurs when populations or groups of populations become isolated from one another so that they are no longer able to exhange genes via sexual reproduction. This might happen when a few individuals disperse to an isolated region (e.g., an island) or when a formerly contiguous range is divided by geological or climatic events (e.g., mountains, rivers, patches of inhospitable environments) that prevent gene flow. If the barrier to gene flow between the two populations remains intact for a long time, the isolated populations will begin evolving independently—a mutation arising in one population lineage will not spread to the other. As a result, the two populations will begin to acquire biological differences.

It should be noted that in this scenario, technically called ***allopatric*** divergence ("allo-" = different; "patria" = homeland), the distinct lineages are initially formed by extrinsic geological events and only later evolve differences as a consequence of being genetically isolated. This is thought to be the predominant mode of lineage splitting. However, in some circumstances lineage splitting can be facilitated by selection for ecological specialization within a population. In such cases, lineage splitting can occur without complete geographic isolation, a phenomenon called ***sympatric*** ("sym-" = same) divergence. For example, sympatric lineage splitting has been proposed for several fish groups, including the cichlids of Lake Victoria. However, because sympatric divergence is less common, we will focus on the allopatric case and how it leads to the production of new evolutionary lineages.

If population isolation is transient, then after the geographic barrier disappears, genes will flow again between the daughter populations, "braiding" them back into a single lineage. However, if the lineages remain isolated, the organisms in the isolated populations will tend to accumulate differences from each other in morphology (physical makeup), physiology, and behavior (Chapter 4). Eventually, these differences may make it impossible for individuals from the two allopatric lineages to mate successfully and/or to create viable offspring with one another. At this point, the separation of the lineages ceases to be dependent on the persistence of an original geographic barrier: reproductive isolation has shifted from being extrinsic, due to geography, to being intrinsic, due to the biological traits of the organisms.

It is a useful simplification at this point to assume that once lineage splitting has been completed, the two descendant lineages will remain isolated

(exceptions are discussed in Chapter 6). This means that, once they have diverged, lineages do not exchange genes by hybridization. This is the underlying reason why evolution can be modeled as a tree rather than as a net. While some groups of organisms, for example, some microbes, do transfer genes between distant relatives, it can still be useful to think in terms of trees. For example, even when the organismic relationships are netlike, the genes themselves will typically have treelike histories (Chapter 6).

Figure 3.6 shows what we might see if we followed the fate of one initial lineage long enough to see it give rise to four living descendant lineages. This example also includes three lineages that were established but then went extinct before the end of the observation period.

In the left panel we have maintained the direction of time from the previous figures, with descendants below their ancestors. When we start looking at longer time frames, it is common to invert the arrow of time, placing the past at the bottom and present at the top. This convention probably arose because older (ancestral) fossils tend to lie in lower strata than fossils of lesser age. Also, the resulting figure, such as that in the right half of Figure 3.6, looks more like a living tree.

These diagrams show simple phylogenetic trees. Because the lines depict actual parent-offspring descent within lineages, they show that at least some of the organisms in the ancestral population are direct lineal ancestors of organisms in the (living) descendant populations. If there are $N + 2$ intervening gen-

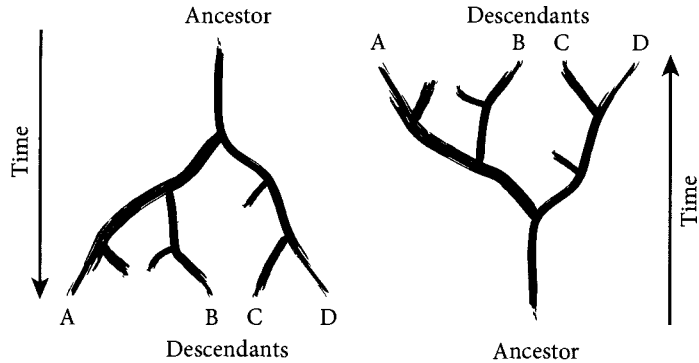


FIGURE 3.6 The branching of an ancestral lineage into four descendants.

erations then you can think of the individuals in the ancestral population as being (great × N) grandparents of the living individuals.

If we were to zoom out as far as we could go from these four living species, we would ultimately find the entire tree of life, whose last common ancestor lived in the truly ancient past, over 3 billion years ago. As with the small tree in Figure 3.6, some long-extinct organisms are direct lineal ancestors of all organisms alive today. While it can be hard to imagine that there are organisms that are the (great × N) grandparents of such different organisms as humans, oak trees, mushrooms, and bacteria, this is what the tree-of-life model implies. As summarized in Chapter 2, there is abundant evidence supporting the tree model and descent from common ancestry.

Except in rare cases, such as laboratory studies of viruses or bacteria, we are not able to watch lineages evolve. Instead of starting from one ancestor and observing evolution occurring in a forward direction, phylogenies are generally approached in the reverse direction. We start from a sample of living tips of the tree and ask, how are they connected back through time? We are effectively taking a cross section of the evolving tree at the present and using information in this time slice to learn about earlier periods of time.

In thinking about evolutionary connections among the tips, it goes without saying that the future of the tree can be ignored. While we should always remember that evolution is ongoing (albeit too slowly to see except in some rapidly evolving groups), trees depict history only. Perhaps less obviously, we do not need to have direct knowledge of any ancestors. Although a tree implies the existence of certain ancestors, and even implies that those ancestors had certain combinations of traits (see Chapter 4), tree thinking is primarily concerned with understanding the evolutionary connections among tips. While ancestors must have existed, we never need to directly interact with ancestors to reconstruct or utilize trees.

The preceding might lead you to wonder about fossils. Are they on the tree and, if so, where are they? While some fossils might be actual ancestors of living species, the best way to approach fossils is to think of them as organisms that were collected in the present (as indeed they were), but stopped evolving a long time ago. That is to say fossils are best viewed as tips of the tree that have a shorter branch (in units of time) connecting them to the (inferred) ancestors. They are treated as living forms that have undergone no evolution in the millions of years since they were entombed in rock.

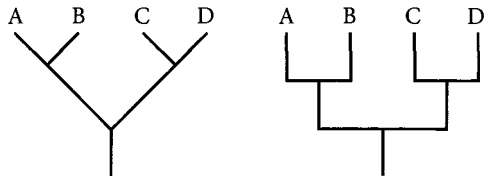


FIGURE 3.7 Two alternative representations of the same four-species tree.

In the case illustrated in Figure 3.6, we will assume that no fossils have been found. Thus, only the four living species are available for study. In this case, all the relevant information in Figure 3.6 can be summarized as either of the simplified tree diagrams shown in Figure 3.7. These show that the initial lineage-splitting event gave rise to two lineages, one of which later split to give rise to descendant species A and B, whereas the other gave rise to C and D. As discussed more fully in Chapter 5, this means that A and B are more closely related to each other than to C and D (and conversely C and D are more closely related to each other than to A and B).

ASEXUAL ORGANISMS

The preceding characterization of phylogenetic trees applies to organisms that reproduce sexually. It is sex, and the resulting potential for gene flow, that glues local populations into cohesive evolutionary lineages. But many different types of organisms reproduce asexually or ***clonally***. So how should we visualize descent relationships in strictly asexual organisms or in a clone of cells?

In asexual organisms each organism has only one parent, which contrasts with sexual organisms in which two parents are required for reproduction. Asexual reproduction, unlike sexual reproduction, is treelike down to the level of individual organisms (or cells within a developing organism). Consider the growth of an asexual aphid population from a single founder arriving on a host plant, as shown in Figure 3.8. If you look closely you will see that it has a perfectly treelike form—lineages split but never merge. We can therefore show the evolutionary history of these aphids in a tree form. Families derived from a single ancestral organism are equivalent to sets of species descended from a single ancestral species. Thus, all the conventions and manipulations described later

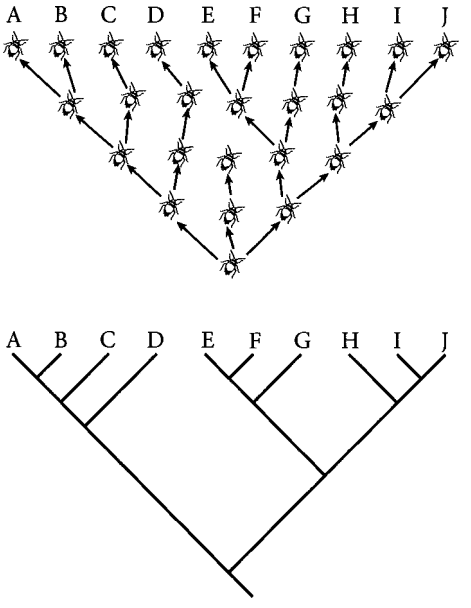


FIGURE 3.8 Treelike history of organisms in asexual lineages, as illustrated with a hypothetical aphid population.

in this chapter can be applied easily and naturally to asexual organisms, right down to the level of individual clones.

TREE TERMINOLOGY AND CONVENTIONS

A tree diagram is made up of lines, called ***branches*** (or ***edges***), connected at nodes. To be considered a tree in the formal sense, the diagram needs to be ***directed***, meaning that time runs in one direction along each branch, and ***acyclic***, meaning that lineages that diverge never subsequently fuse. Figure 3.9 shows a simple rooted tree with some of its parts labeled. The version on the left is in a rectangular format, whereas the tree on the right is in a diagonal format.

In an evolutionary context, the labels at the top of a tree could be individual species, individual organisms that represent particular species, or sets of related

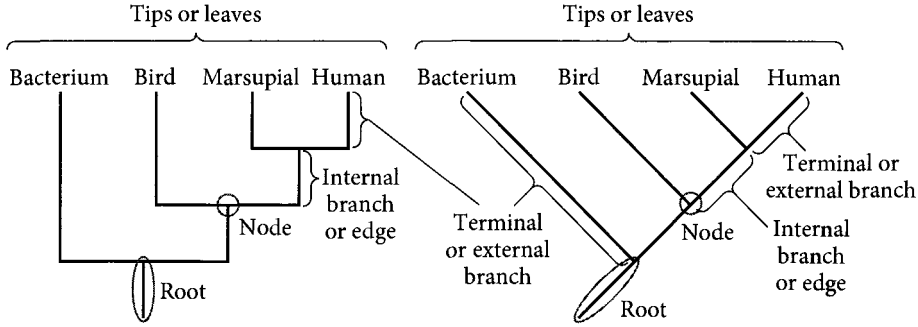


FIGURE 3.9 Terms associated with phylogenetic trees.

species that constitute one branch of the tree of life. In some situations they could be individual genes. The most common terms for the items represented by these labels are ***tips*** or ***leaves***, but you might also see them called ***terminals***. If the tips have scientific names, they may be called taxa (singular = taxon). The branches represent evolving lineages, whereas ***nodes*** correspond to lineage-splitting events. A node marks the last common ancestor of organisms in the daughter lineages. Whereas the ***internal branches*** (or ***internodes***) connect two nodes, ***external branches*** connect a tip and a node. The ***root*** of the tree is a special node that marks the point where time enters the diagram. The root is usually indicated by an external branch whose tip is unlabeled—generally drawn on the opposite side of the diagram to the labeled tips.

When describing trees, it can be useful to have a way to refer to a piece of a tree that is descended from one particular ancestral lineage. A ***clade*** is a piece of a phylogeny that includes an ancestral lineage and all the descendants of that ancestral lineage. Alternatively, just focusing on living taxa, a clade can be defined as a set of tips that comprise all the living descendants of one particular ancestral node. Clades have the property of ***monophyly*** (from the Greek for "single clan") and, thus, may also be called monophyletic groups. As shown in Figure 3.10, a clade or monophyletic group is easy to identify visually: it is simply a piece of a larger tree that can be cut away from the rooted part of the tree with a single cut. If one needs to cut the tree in two places to extract a set of tips, then that group is non-monophyletic and is not a clade. See Chapter 5 for further discussion of non-monophyletic groups.

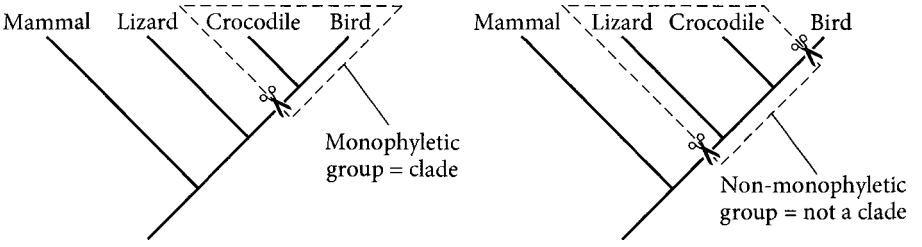


FIGURE 3.10 Distinguishing a monophyletic group (or clade) from a non-monophyletic group. Monophyletic groups can be separated from the root by a single cut, whereas separating non-monophyletic groups requires at least two cuts.

By analogy to family trees, we may refer to the two descendants of a single node as ***sister groups*** or ***sister taxa***. This convention provides a useful way to verbally describe a tree topology. For example, in Fig. 3.10, bird and crocodile are sister taxa and the lizard lineage is the sister taxon to the bird+crocodile clade. Note that the sister taxon relationship is unique; a taxon can have one and only one sister taxon. If an ancestral lineage branched simultaneously into three or more daughter lineages (discussed further below), then the daughter lineages would not have sisters.

TREE TOPOLOGY

The most basic information in a tree is the relative branching order, or ***topology***. Tree topology, that is, which lineages lead to which tips, is an important predictor of the distribution of traits among tips (Chapter 4). Tree topology also tells you which organisms are more or less closely related to each other (Chapter 5). Here, we will focus on how to correctly read tree topology because many students find this challenging. Later in this chapter we will discuss tree diagrams that include information on the amount of evolution occurring on a branch and/or the duration of branches.

If you recall the way that a phylogeny "grows" by ancestral lineages splitting, it is arbitrary which descendant lineage is shown on each side of the figure. Trees that are topologically equivalent can look quite different when different branches are positioned to the right or left. We can "spin" parts of a tree around

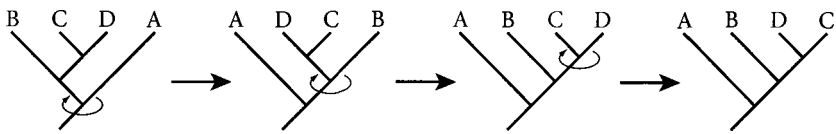


FIGURE 3.11 **Rotating branches at nodes does not alter relationships.** The trees can be interconverted by rotating the indicated branch.

any internal branch without changing the topology. So long as you can get from one tree to another by rotating nodes or reshaping branches, but not cutting and reattaching branches, those trees have the same topology. For example, Figure 3.11 shows four equivalent trees with an indication of which nodes need to be rotated to get from one to the other. In each tree, (C, D) is a clade and so is (B, C, D).

Notice in Figure 3.11 that in each tree the tips are ordered differently despite having the same topology. This shows us that the left-to-right ordering of tip labels is arbitrary and should not be used to extract information from a tree diagram. In particular, the ordering of tips should never be taken to convey information about evolutionary "advancement" (Chapter 2).

For many people the kinds of mental gymnastics needed to see the equivalence of different tree topologies is challenging. These skills can be developed by playing with computer programs that allow branches to be moved around graphically (e.g., Mesquite) or by manipulating physical models of trees (e.g., made from pipe cleaners). Nonetheless, it is helpful to also know about some formal rules that can be applied to determine if two trees have the same topology.

One method is to imagine that the lineages of a tree are a set of paths with signposts at each junction, indicating the tips (villages, if you will) that are served by each alternative route. If you walked up from the root on the first tree in Figure 3.11, the first junction you would come to would have one sign pointing to village A and one pointing to B, C, and D (see Figure 3.12). We can write this out in the ***splits*** format: A|BCD. All that matters in this convention is which villages are clustered on each side of the vertical line. Thus, A|BCD is the same as A|DCB, BCD|A, DBC|A, and so on.

If you then walked up the BCD path, you would come to a B|CD signpost (Figure 3.12). The tree also has a C|D split, but this does not add any information that was not given by the B|CD split. You already knew that there was a road leading to C and D, so there must be a C|D signpost. Therefore, the pertinent information in the tree is summarized by the two splits: A|BCD and

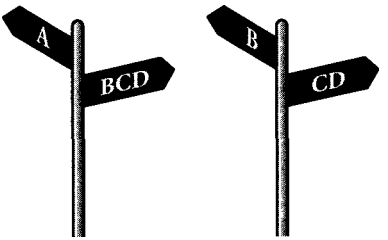


FIGURE 3.12 Trees as road signs. Branching points of trees are like forks in the path of evolution, with either side leading to a different clade or tip. The figure shows the two road signs or "splits" that are shared by all the trees in Figure 3.11.

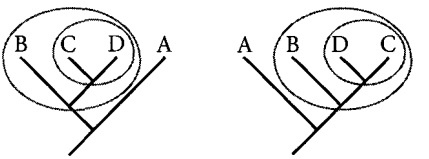


FIGURE 3.13 Comparing trees by clade composition. The two clades (CD) and (BCD) are the same, showing that these two trees are equivalent.

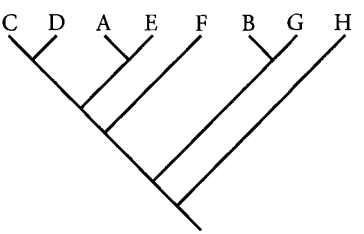


FIGURE 3.14 An eight-taxon tree.

B|CD. If you do the same exercise on the other trees in Figure 3.11, you will see that each has the same two splits, showing that they all have the same topology.

A slightly different method is to list clades: sets of tips that are descended from a particular internal node. For example, because the two trees in Figure 3.13 contain the same two clades, (BCD) and (CD), we can see that they have the same topology.

The clade convention is not only useful for seeing if two trees are equivalent, it also provides a simple way of writing a tree topology in text format, by listing all the clades within parentheses. For example, the trees in Figure 3.13 can be rewritten in so-called Newick format as (A(B(CD))). This convention can be scaled up to an indefinite number of tips. For a slightly more complex example, the tree in Figure 3.14 can be written as (H((GB)(F((EA)(DC))))). This tree description is not easy to read by humans, but is a standard way to input trees into computer programs.

DIFFERENT TREE STYLES

There are several alternative styles in which trees can be drawn (Figure 3.15). Most trees drawn so far in this chapter have been in the ***diagonal*** or ***rectangular*** format and have had the root at the bottom so that time points up. These same formats are sometimes used in a different orientation. For example, Figure 3.15 shows the same topology drawn in a diagonal-up, diagonal-down, and rectangular-right format. This figure also shows one additional tree style, the circle tree. A ***circle tree*** has only one orientation: time always runs outward from the middle.

The four trees in Figure 3.15 are all equivalent to one another. This can be established using the clade or signpost methods or using mental gymnastics to convert one to the other by twisting and bending branches. In all cases, the tree shown has the topology (A(B(C,D))).

The choice among different tree formats is guided by practical and stylistic issues rather than biological factors. Diagonal trees are efficient because few lines need to be drawn. Psychological research has shown that diagonal trees may confuse students, who sometimes misinterpret the long diagonal line from the root to taxon D as indicating that taxa A, B, and C descended "from D." This confusion is easily overcome by recalling the meaning of a tree diagram and by becoming fluent in converting topologies from a diagonal to a rectangular format.

Rectangular trees are tidy and provide horizontal lines on which to insert text. Circle trees are useful when one wishes to include many tips in a compact

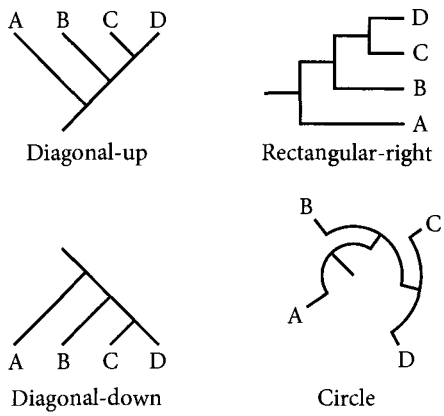


FIGURE 3.15 Four alternative representations of the same topology.

space. They also have the more subtle advantage of not being easily misinterpreted as implying that evolution was headed toward one privileged tip. However, because they are tricky to draw and take practice to read, they are used only when the other formats are incompatible with the available space.

Because you may encounter a diversity of tree formats, it is important to be able to tell whether trees in different formats contain the same information. In this book we intentionally use a mixture of tree formats to give you practice in working with these different forms.

MERGING AND PRUNING

The entire tree of life is very, very big, including several million known living species. A tree depicting the relationships among a single representative of every species would be bewilderingly large. Furthermore, within a single named species, multiple tips can be recognized: subspecies, populations, or individual organisms. Also, there is no obvious limit to the number of fossils we might eventually discover, and each fossil form is best considered a tip, as discussed earlier in this chapter. In light of the immensity of the grand tree of life, the tree metaphor has utility only because it is resilient to certain simplifications—resilient in the sense that statements based on a small piece of the tree will be true for the tree as a whole. Two ways of simplifying trees are especially important: ***pruning*** and ***merging***.

Phylogenetic trees only depict relationships among the terminals that are included in the diagram. Nonetheless, the treelike form has the desirable property that pruning tips off a tree does not change the relationships of the remaining tips. For example, given the tree on the left in Figure 3.16, the pruned tree on the right correctly represents the phylogeny for the remaining species.

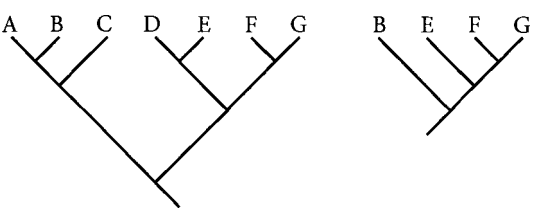


FIGURE 3.16 Pruning does not alter relationships of the included tips. Removing three tips (A, C, D) from the tree does not change the relationships among the remaining taxa (B, E, F, G).

All we have done is cut off three tips (A, C, and D) and then straightened the remaining branches. Two trees are said to be ***compatible*** if there is one larger tree topology that can be converted to either of the two trees by selective pruning of branches (one of the trees can be identical to the larger tree).

The resilience of trees to pruning is an important feature that explains why they are such good devices for communicating information. Because of this property it is possible to make accurate statements of evolutionary kinship without having to list every species that ever lived. Or, conversely, adding a newly discovered species to a well-established tree has no effect on the relationships among the species that were already included.

Stability in the face of pruning takes advantage of the fact that each tip is connected to the rest of the tree (and hence to all other tips) by only one connection. This can be illustrated by imagining a literal tree (Figure 3.17). A squirrel seeking to climb from tip A to tip Z along branches (without jumping) follows exactly the same path regardless of how many other tips and branches have been pruned off. This is a manifestation of the acyclic nature of a tree.

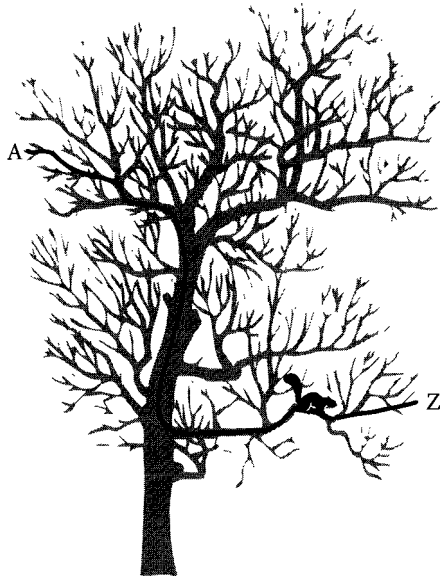


FIGURE 3.17 The path of a squirrel from leaf A to leaf Z is the same regardless of how many other branches are on the tree.

Figure 3.18 provides an illustration of tree pruning. In the upper panel we show the tree that reflects the currently accepted relationships among these mammals (all except the tenrec are in a clade called Afrotheria), and we have pruned off several tips to yield the smaller tree to the right. In the lower panel, we have taken the same "full" tree and have removed a different set of branches. While the two pruned trees may look different, they have the same topology for the species that are included. The three trees are compatible with one another.

Determining by eye whether a small tree is a validly pruned version of a larger tree takes practice. A modification of the clade method for testing

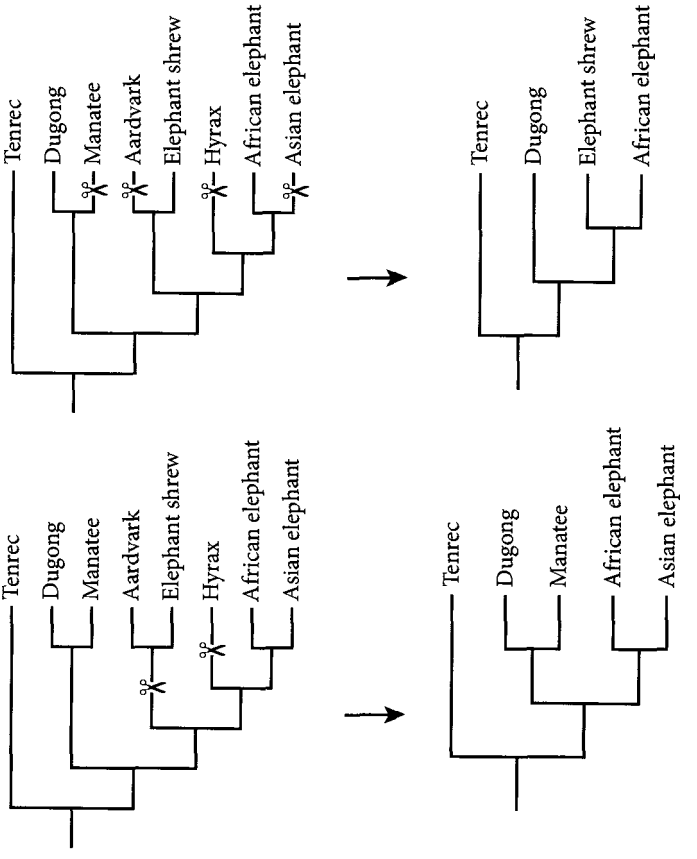


FIGURE 3.18 Two pruned versions of the same larger tree.

topological identity can be used to see if trees with different numbers of tips are compatible. The rule can be broken down into the following steps: (1) Find a clade on the smaller tree and note the tips that are included. (2) Find the smallest clade on the larger tree that includes all those tips. (3) Note any tips that are in the clade on the larger tree that were not in the clade on the smaller tree. (4) If these extra tips occur anywhere on the smaller tree, then the two trees are incompatible: the smaller tree is not a pruned version of the larger tree. (5) If, after considering all clades in the smaller tree, you do not find any cases of incompatibility, then the smaller tree is a validly pruned version of the larger tree.

For example, the upper pruned tree in Figure 3.18 has a clade that includes just African elephants and elephant shrews. The smallest clade on the unpruned tree that includes these two tips also includes the aardvark, hyrax, and Asian elephant tips. However, since none of these three taxa are anywhere on the pruned tree, we can conclude that the African elephant + elephant shrew clade is compatible with the full tree. Repeating this procedure for all the clades can show that the pruned trees in Figure 3.18 are compatible with the full trees to the left.

In addition to being resilient to pruning, trees can also be simplified by merging a clade into a single tip. Regardless of how large a clade is merged, the basic topology of the tree remains the same. For example, instead of displaying every species in clade H in the tree on the left in Figure 3.19, we can redraw the tree with 'H' merged into a single terminal.

It is important to understand that merging is a valid maneuver for clades but not for non-monophyletic groups. To see why, consider a couple of examples using the tree in Figure 3.19. First, imagine that you renamed the non-monophyletic group C + D + E as "I." Wherever you placed I on the merged tree would imply an incorrect placement for at least one tip. For example, if "I"

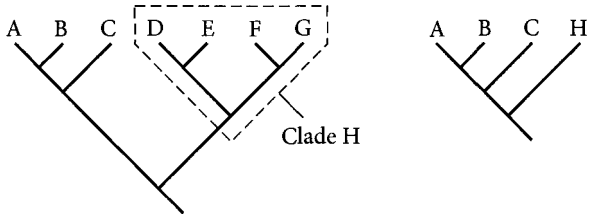


FIGURE 3.19 Merging a clade into a single tip. The two trees are identical given that taxa D–G are merged into clade H.

were placed as the sister taxon to F + G, it would incorrectly suggest that C is more closely related to F + G than it is to A + B. This is a major reason why the current convention is to give formal taxonomic names only to monophyletic groups of organisms (Chapter 5).

THE TIME AXIS

A rooted tree follows the fate of one ancestral lineage through a series of lineage branching events, usually leading to a set of living species. The tree is a historical chronicle: the nodes and branches represent ancestral populations that lived at some particular time in the past. A tree diagram must, therefore, contain some implicit information on the relative timing of different lineage-splitting events. However, you should be careful not to read too much temporal information into a tree diagram.

Two nodes that are on the same path from the root have a fixed relationship to one another. The node closer to the root represents a population of organisms that is ancestral to the other node, and therefore lived earlier. For example, in Figure 3.20, node *b* is a descendant of node *a*. The latter must represent a population of organisms that lived after the former. Node *c* is also a descendant of node *a*, and must, likewise, have lived after node *a*. To make this easier to see, the figure on the right adds arrowheads pointing along lineages from ancestors to descendants.

While the tree contains information about the relative ages of ancestral and descendant nodes, this diagram does not contain information about the relative ages of nodes that are not on the same path from the root. For example, Figure

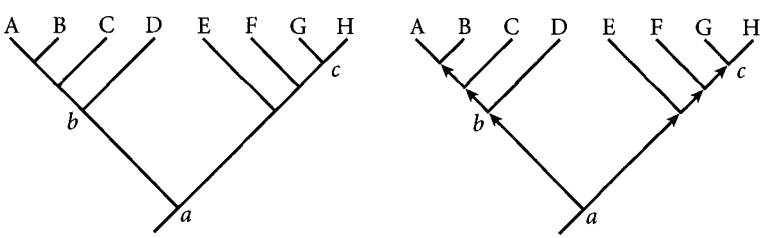


FIGURE 3.20 Trees contain information about the relative age of nodes that are on the same path from the root. We can infer that nodes *b* and *c* are both younger than node *a*, because they are both descendants of node *a* (as indicated by the arrows in the right panel). However, without explicit temporal information, we cannot determine the relative ages of nodes *b* and *c*.

3.20 does not contain information on the relative ages of nodes *b* and *c*. It might be tempting to infer that node *c* lived after node *b* because there are two nodes between *a* and *c* but no intervening nodes between *a* and *b*. This reasoning is flawed because this diagram just shows tree topology. There is no information about branch duration in this diagram. For example, the three branches between nodes *a* and *c* could each represent a short period of time, summing to less total time than the single internal branch between nodes *a* and *b*.

A convenient way to summarize the ages of nodes (when known) is to draw branch lengths proportional to time, usually with an associated scale to allow one to read off the estimated age of an internal node. Such diagrams are called ***chronograms***, because they contain information on time, as contrasted with ***cladograms***, which only depict topology and clade membership. Figure 3.21 shows a chronogram that matches the cladogram in Figure 3.20. It is assumed that all the tips shown are extant (i.e., still living), meaning that they lived zero millions of years ago. By dropping a line from internal nodes to the timescale, we can see that node *c* is older than node *b*. Chapter 11 introduces molecular dating methods that may be used for constructing chronograms.

An intermediate situation between a cladogram and a full chronogram is encountered when certain nodes or tips within a tree are assigned ages, but branch lengths are not drawn proportional to time. Let us start with a case where certain tips are fossils of known age, as shown in Figure 3.22. Given that tip F is a fossil that is dated at 55 Ma, what else can we infer?

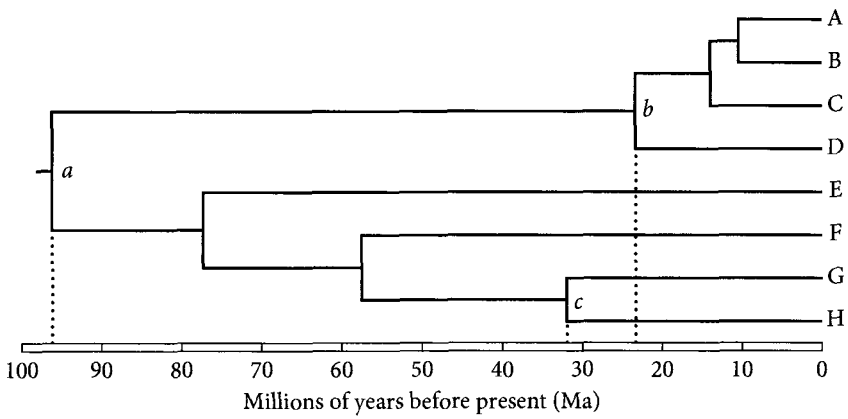


FIGURE 3.21 A chronogram showing the timing of the branching events. This figure shows that node *c* predates node *b*, something that could not be inferred from the cladogram depicted in Figure 3.20.

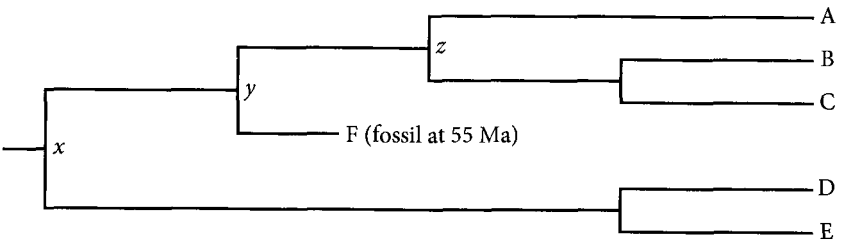


FIGURE 3.22 Using a dated fossil to place limits on nodal ages. Given that F is dated at 55 Ma, we can infer that nodes *x* and *y* are both at least 55 Ma. We cannot, however, constrain the age of node *z* or any of the unlabeled nodes without making additional assumptions.

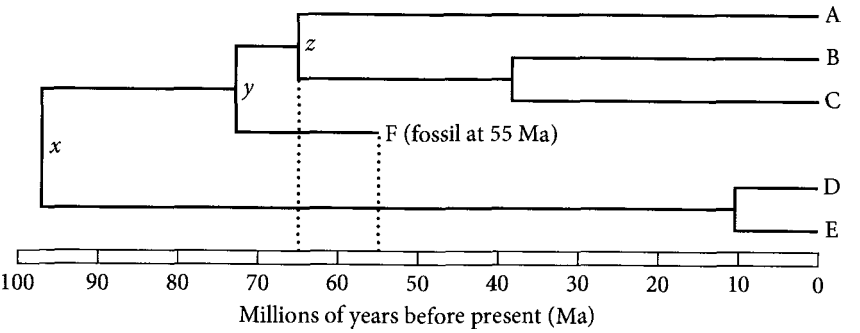


FIGURE 3.23 A chronogram showing that fossil F lived after node *z*.

Because nodes *x* and *y* are ancestral to F, it is valid to assume that they existed at least 55 Ma. It might be tempting to infer additionally that node *z* is younger than 55 Ma, based on the reasoning that it occurred after the origin of the F lineage. However, such reasoning would be invalid. The branch between node *y* and F could be of long duration, while the one leading from node *y* to node *z* could be of short duration. In that case node *z* could predate 55 Ma, as shown in Figure 3.23. Thus, based only on Figure 3.22, we have no direct information on the age of node *z* or either of the other two unmarked nodes.

BRANCH LENGTH AND EVOLUTIONARY RATE

While topology suffices for some purposes, such as delimiting taxonomic groups or inferring evolutionary relationships (Chapter 5), some downstream

uses of trees require information on the length of branches, where length usually represents the relative probability that a character would change state on a particular branch. For example, if you are quantitatively analyzing patterns of trait evolution (Chapter 10), knowing that more evolutionary changes tend to occur on some branches than others can have a substantial effect on your conclusions. In many cases, expressing branch length in units of time (i.e., in the form of a chronogram) is all that is needed. In other cases, it can be useful to draw branches such that their length is proportional to the amount of evolution that is inferred to have occurred on the branch, most often expressed as the average number of changes occurring to each character used in a particular analysis (Chapter 8). Trees with branch lengths drawn proportional to the amount of evolution are called ***phylograms***.

Figure 3.24 shows a sample phylogram from a study of the evolution of cotton (*Gossypium*) and its wild relatives. The scale bar at the bottom of the diagram indicates that the branch lengths are proportional to the amount of evolutionary change. Phylograms are among the most common tree diagrams in research literature, but they are less common in secondary literature and in textbooks, where cladograms and chronograms predominate.

The branch lengths on a phylogram relate to a specific set of traits, most commonly the gene sequences that were used to infer the tree (Chapters 7–8). For example, if we were considering the evolution of the hemoglobin protein, then length might be drawn proportional to an estimate of the proportion of amino acid sites that changed on each branch. In the case of Figure 3.24, the branches are drawn proportional to the number of substitutions estimated to have occurred at each site in a portion of the plastid (i.e., the chloroplast) genome of these plants. The length of a branch is its duration multiplied by its average rate of evolution. Because the rate of evolution can vary across branches, phylograms are not the same as chronograms.

If two sister lineages are of different lengths, and if both have living representatives, then the rate of evolution must have differed between them. The longer branch has accumulated more changes and, thus, has evolved at a faster rate than the short branch. On Figure 3.24, *Gossypium* and *Hibiscus* are part of a clade called the Eumalvoideae with long branches, whereas the baobab, *Adansonia*, is in a clade called Bombacoideae with relatively short branches. How should this be interpreted? All one can safely assume is that for *this gene* the rate of molecular evolution was higher in Eumalvoideae than in Bombacoideae. It could be the case that all genes evolved more rapidly in Eumalvoideae

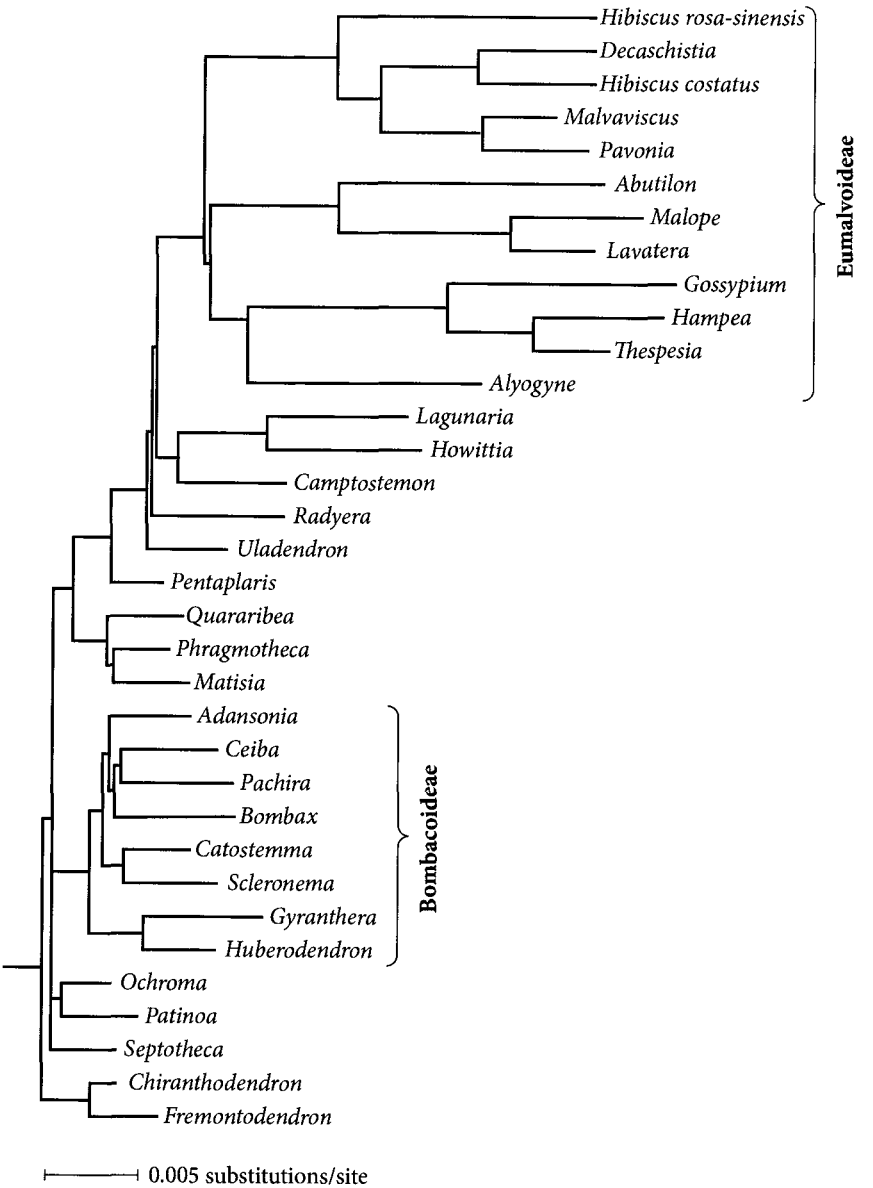


FIGURE 3.24 **Example of a phylogram.** The scale bar indicates the relationship between the branch lengths shown and the average number of substitutions that occurred at each site in the sequence. Adapted from Baum et al. (2004).

than in Bombacoideae, but this conclusion should not be drawn from a single phylogram.

COMMUNICATING PHYLOGENETIC UNCERTAINTY

Up until this point, all the trees we have shown were ***binary***: ancestral lineages split into just two descendant lineages. The splits are ***dichotomous*** (Greek for "cutting into two"). A fully binary tree is also called ***fully resolved***. In parts of the tree of life where lineage splitting is a rare event, it is probably reasonable to assume that all lineage-splitting events are dichotomous. But it is easy enough to imagine cases in which an ancestral lineage splits more or less simultaneously into multiple descendants. For example, a widespread species might become fragmented into multiple isolated populations as a result of a change in the climate. If several of these populations persisted to establish new lineages, the result would be a node with more than two descendant lineages, a ***polytomy***. Figure 3.25 compares binary and polytomous trees drawn in either diagonal or rectangular format.

It is certainly possible that true phylogenies have polytomous nodes, so-called ***hard polytomies***. More commonly, polytomies in tree diagrams indicate uncertainty as to the correct branching pattern. Recall that the trees that appear in research publications, textbooks, or websites are inferred from data: they are hypotheses of actual evolutionary relationships. Thus, even if the true tree were fully binary, the data might be insufficient to resolve all the relationships. Polytomies that are used to communicate uncertainty in the tree topology (rather

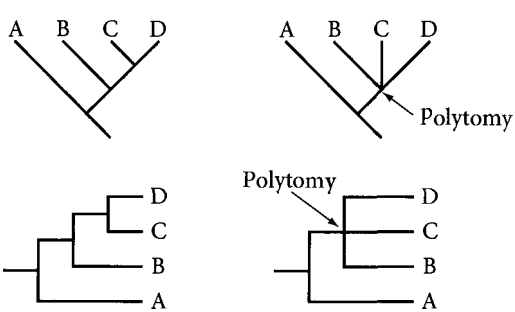


FIGURE 3.25 Polytomies express uncertainty in phylogenetic relationships. The clade containing B, C, and D is collapsed into a polytomy in the trees on the right.

than ancestral lineages actually splitting into multiple descendants) are called ***soft polytomies***.

Consider a study that has ruled out most of the possible trees, but cannot rule out the two trees in Figure 3.26. The remaining uncertainty can be captured with a ***consensus tree*** using the two possible trees as ***input trees***. The consensus tree shown in Figure 3.26 is a ***strict consensus tree***: a tree composed only of clades that occur on all input trees. Both input trees include the clades (ABC), (DEFG), and (FG), and therefore the consensus tree includes just these three resolved clades. Internal branches that are not present on all the input trees are collapsed into a polytomy. The clade (DE), for example, occurs only in one of the input trees and is therefore not shown on the consensus tree. There are other kinds of consensus trees, but we will not describe them here.

You may be wondering how to interpret polytomies if they can represent either an ancestral lineage splitting simultaneously into multiple descendant lineages or phylogenetic uncertainty. The safest interpretation is to view the polytomy as an indication of uncertainty, where that uncertainty includes the possibility of a hard polytomy.

Polytomies in tree diagrams represent complete uncertainty. However, it is common for an analysis of real data to yield a tree whose branches receive different levels of support by those data. Phylogeneticists have developed a

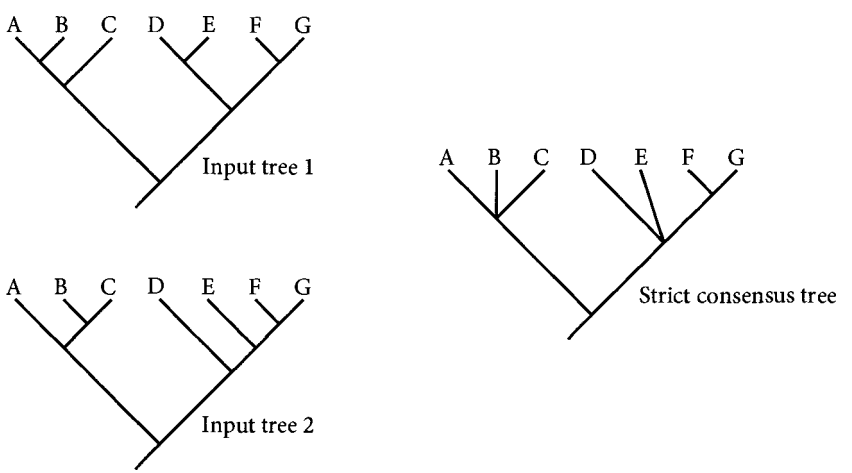


FIGURE 3.26 Combining two resolved input trees into a strict consensus tree with polytomies. The consensus tree contains only clades that are present in both input trees.

number of different ways to annotate a tree to indicate the degree of confidence that should be associated with each clade. The strength of support for a particular clade is typically indicated by placing a number on the branch subtending that clade. The most commonly used measures are ***bootstrap percentages*** (also called ***bootstrap scores***), which range from 0 to 100% and ***posterior probabilities*** (also called ***clade credibilities***), which range from 0 to 1.0. The meaning of these numbers is explained in Chapters 8 and 9. For now, it will suffice to know that the higher the number, the more strongly the data support the clade descended from the annotated branch (or node). While the thresholds are subjective, a rough rule of thumb is that clades with bootstrap scores greater than 80% or posterior probabilities greater than 0.95 are considered well supported.

As an example, consider Figure 3.27, which shows part of a phylogram from a scientific paper (Medina et al. 2001). While most of the sampled tips

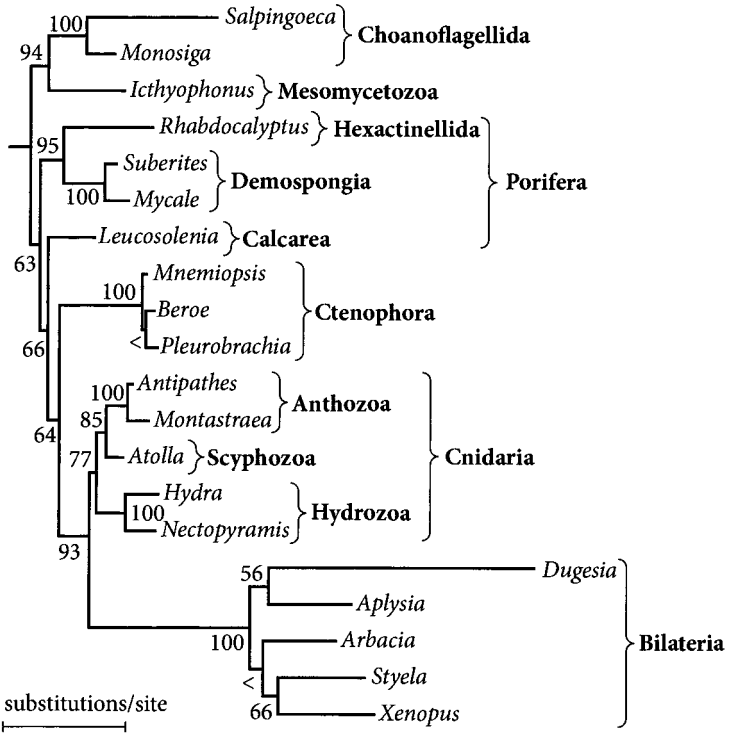


FIGURE 3.27 A phylogeny of animals and their closest relatives. Numbers are bootstrap scores (scores <50% are marked "<"). Scale bar as in Figure 3.24. Adapted from Medina et al. (2001).

may be unfamiliar, you may know a number of the major clades represented. Each branch of the tree has an associated bootstrap score. Values less than 50% are typically not provided. Knowing that all internal branches have bootstrap scores, and that these scores refer to the clade descended from that branch, you can see which number refers to which clade. This allows us to determine which results were strongly supported by this analysis. For example, Figure 3.27 provides quite strong support (93%) for a clade comprising both Bilateria (the clade that includes the vast majority of living animal species, including us) and Cnidaria (e.g., jellyfish and corals). In contrast, this study provides relatively weak support (77%) for the monophyly of Cnidaria. Likewise, while this tree contradicts the monophyly of sponges, it does so only weakly; this is because the support for Calcarea forming a clade with other animals rather than the remaining sponges is only 66%. While it can take practice to read trees in this manner, an ability to do so opens a wonderfully rich array of scientific literature that summarizes phylogenetic data using these conventions.

UNROOTED TREES

Rooted trees contain information about the flow of time, which allows us to discern the pattern of descent from common ancestry and the direction of trait evolution (Chapter 4). Rooted trees are therefore essential for most downstream uses of phylogenies. While it is only necessary to understand rooted trees in order to read much of the secondary phylogenetic literature, you will need some understanding of ***unrooted trees*** and how trees are rooted if you plan to read the primary phylogenetic literature. Additionally, developing a clear sense of how rooted and unrooted trees differ can help you to achieve a deeper understanding of trees in general.

An unrooted tree is a tree without a defined root. In an unrooted tree the branches represent evolutionary lineages, but unlike a rooted tree, we do not know which way evolution preceded along the lineage. Because a clade comprises an ancestor and all its descendants, we need temporal information to identify clades. Thus, the internal branches of an unrooted tree do not denote clades but rather split the taxa into two sets of lineages that are attached (directly or indirectly) to the two ends of the branch.

As an example, consider a rooted tree for selected archosaurs (Figure 3.28a). Figure 3.28b shows an unrooted version, obtained by removing the root and collapsing the lowermost internal branch into a polytomy. The figure may

seem to imply that the true root is on the crocodile lineage or between a crocodile + pterosaur clade and the dinosaurs, but this cannot be assumed. Because this is an unrooted tree, in the absence of extra information, we should be open to the possibility that the true root lies along any branch of the tree.

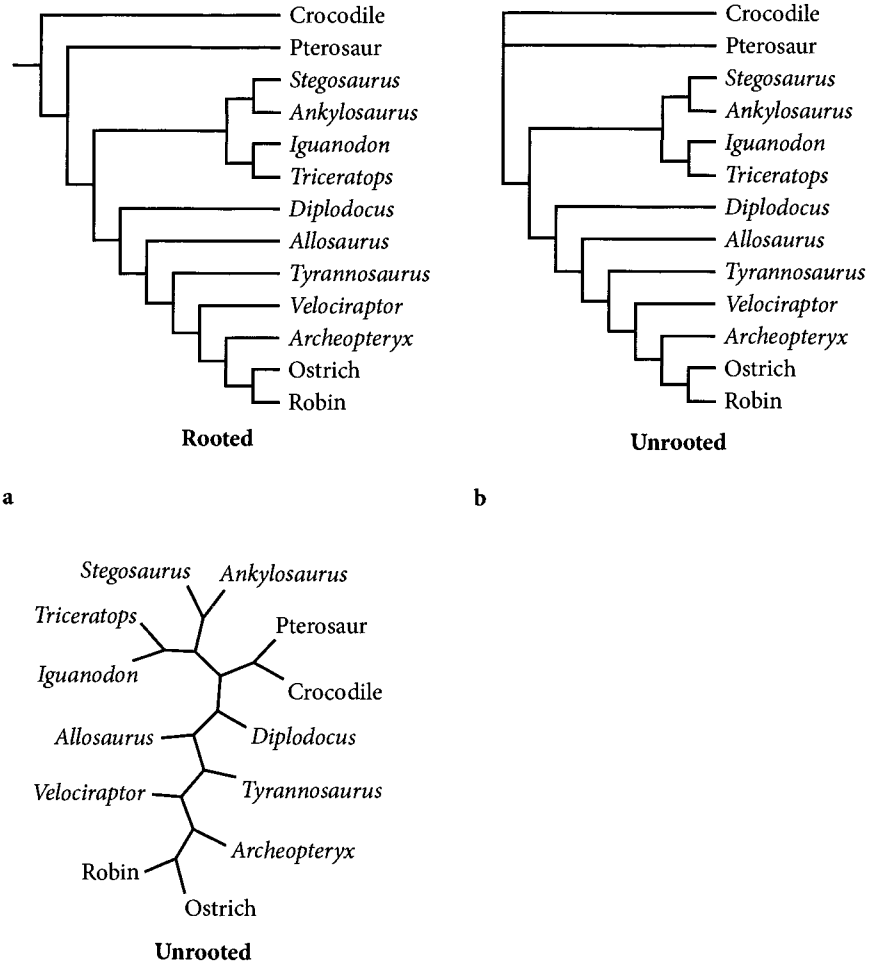


FIGURE 3.28 The same tree topology in rooted (a) and two different unrooted (b, c) tree formats.

To avoid implying a root, the tree has been redrawn in a spread-out style (Figure 3.28c). Like the other trees in Figure 3.28, this tree is binary in that each node has three branches, one corresponding to the ancestral lineage and two to descendant lineages. However, because the trees in Figure 3.28b and c are unrooted, we cannot tell which lineages are ancestral and which are descendant.

To see that the three trees in Figure 3.28 are topologically identical, confirm for yourself that to get from one to the other you need only remove the short branch leading to the root and then unbend, resize, and rotate branches. No additional branches beyond the root branch have to be cut. Once again, imagine that the trees are made of pipe cleaners (but whose length can change) and you can rearrange the first tree to yield the second or the third.

If this physical modeling is difficult, try the approach of listing clades (see earlier section on Tree Topology) to establish the topological identity of these three trees. First list the clades in the rooted tree. Then list the taxa on the unrooted tree that are separated from each other by an internal branch, using a vertical line to indicate which taxa are on which side of the internal branch. These are splits, also called ***bipartitions***. For example, as shown in Figure 3.29, there is one internal branch that divides *Stegosaurus, Ankylosaurus, Triceratops, Iguanodon, Pterosaurus,* and crocodile on the one side from *Diplodocus,*

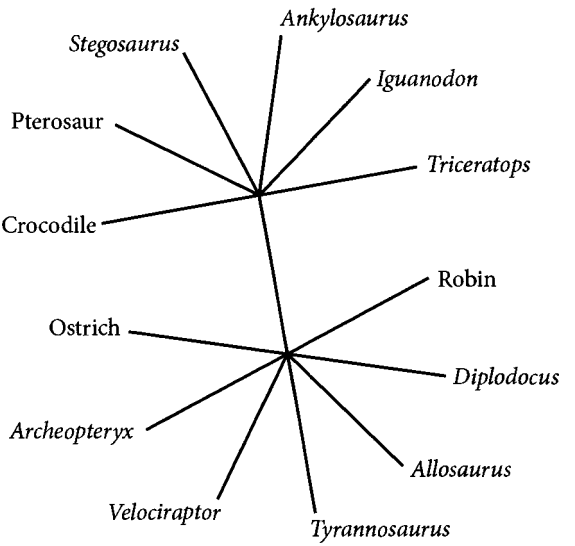


FIGURE 3.29 One of the splits (a bipartition) in the archosaur tree (Figure 3.28).

Allosaurus, Tyrannosaurus, Velociraptor, Archaeopteryx, ostrich, and robin on the other. This could be written as: *Stegosaurus, Ankylosaurus, Iguanodon, Tricerotops, Pterosaur*, crocodile | other species. Table 3.1 lists all the splits seen in Figure 3.28b and c. For each split you need to look at the rooted tree (Figure 3.28a) and ask, Does the set of taxa before the line or the set of taxa after the line or both correspond to a clade on the rooted tree? If the answer is yes for each split on the unrooted tree, and if all clades on the rooted tree correspond to one of the splits, then the rooted and unrooted trees match.

Assuming that you know where the root should go on an unrooted tree, you can easily convert an unrooted tree back into a rooted tree. Rooting an unrooted tree just involves adding an additional node to one of the branches and reorienting the tree relative to that node. Figure 3.30 illustrates three ways to root the same unrooted tree.

You may wonder, How do I decide where to place the root on an unrooted tree? When reconstructing trees (Chapters 7 and 8), scientists generally use one of two methods to decide on how to root the trees. Most commonly, an analysis will include a group that is known to be outside of the group whose relationships are being studied: an ***outgroup***. When an unrooted tree is obtained from a study, the root is placed between the ingroup and the outgroup. Alternatively, in some situations, for example, when a ***molecular clock*** applies (Chapter 11), we can infer the position of the root based on the relative lengths of different branches.

TABLE 3.1 List of splits in Figure 3.28b and c

Stegosaurus, Ankylosaurus | other species
Iguanodon, Triceratops | other species
Stegosaurus, Ankylosaurus, Iguanodon, Triceratops | other species
Pterosaur, Crocodile | other species
Stegosaurus, Ankylosaurus, Iguanodon, Triceratops, Pterosaur, Crocodile | other species
Allosaurus, Tyrannosaurus, Velociraptor, Archaeopteryx, Ostrich, Robin | other species
Tyrannosaurus, Velociraptor, Archaeopteryx, Ostrich, Robin | other species
Velociraptor, Archaeopteryx, Ostrich, Robin | other species
Archaeopteryx, Ostrich, Robin | other species
Ostrich, Robin | other species

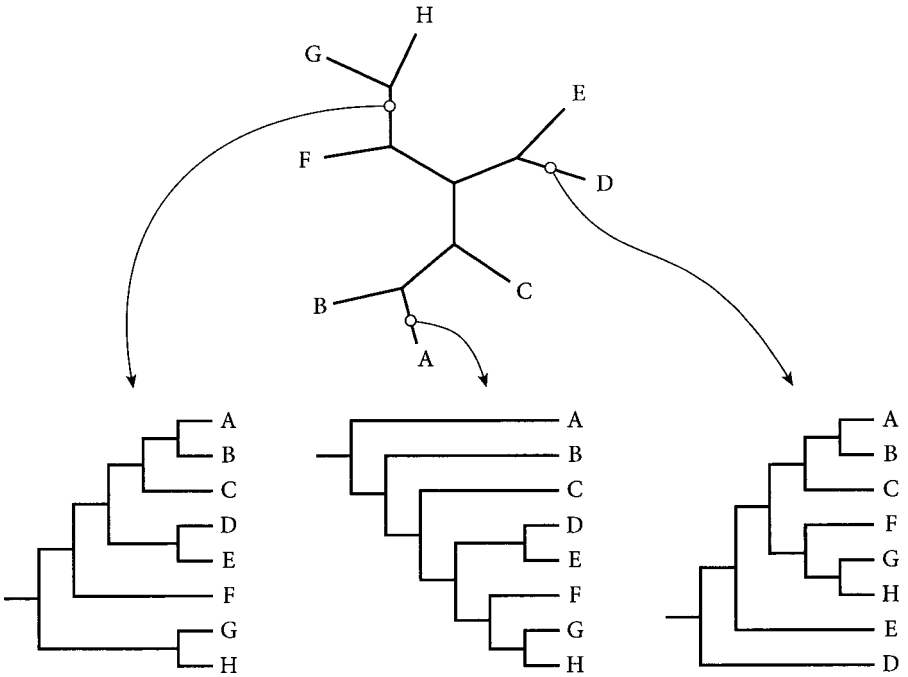


FIGURE 3.30 Three alternative ways to root the same unrooted tree.

TREE-TO-TREE DISTANCES

When working with two trees that are not identical, it may be useful to determine how similar they are to one another. For a number of purposes, it is valuable to quantify this as the degree of difference, that is, the ***tree-to-tree distance.*** It is possible to quantify tree-to-tree distance while taking into account both topology and branch lengths. However, since branch lengths make the analysis significantly more complicated, we will only introduce methods that consider tree topology while ignoring branch lengths.

There are several methods to measure the distance between two tree topologies, of which we will mention two. For simplicity we will only consider trees that are fully resolved (see earlier section on Communicating Phylogenetic Uncertainty), although these basic approaches can be adapted to handle polytomous trees. The first way to quantify the distance between two tree topologies

is to count the proportion of shared clades. For example, Figure 3.31 shows three trees, each with seven clades. Tree 1 has two clades in common with tree 2, (ABCDEFGH) and (AB), and five discordant clades. Tree 1 has three clades in common with tree 3, (AB), (GH), and (ABCDEFGH), and four discordant clades. The number of discordant clades is a measure of distance: tree 1 has a distance of 5 to tree 2 and a distance of 4 to tree 3. This suggests that tree 1 is more similar to tree 3 than to tree 2.

A problem with counting clades as a way to measure the distance between trees is that simple rearrangements can disrupt many clades simultaneously. If you compare trees 1 and 2 more closely, you will see that the entire difference between them is due to the movement of one clade, (AB), which is a sister group to taxon C in tree 1 and to taxon G in tree 2. This suggests that a more appropriate way to measure the distance between two topologies is to count the number of tree rearrangements needed to convert one topology into another.

There are several tree rearrangement methods, of which we will introduce one: ***subtree pruning and regrafting***, or SPR. As the name indicates, this maneuver entails cutting a piece off the tree and reattaching it in a new location. The word ***subtree*** rather than clade is used because single tips can be pruned and regrafted. Also, since SPR rearrangements are usually applied to unrooted trees (Figure 3.32), it is unclear which of the two subtrees is a clade.

While it may be difficult to calculate for large trees, it is usually possible to determine the minimum number of SPR rearrangements needed to convert one specific tree topology into another. This can be used as a measure of tree

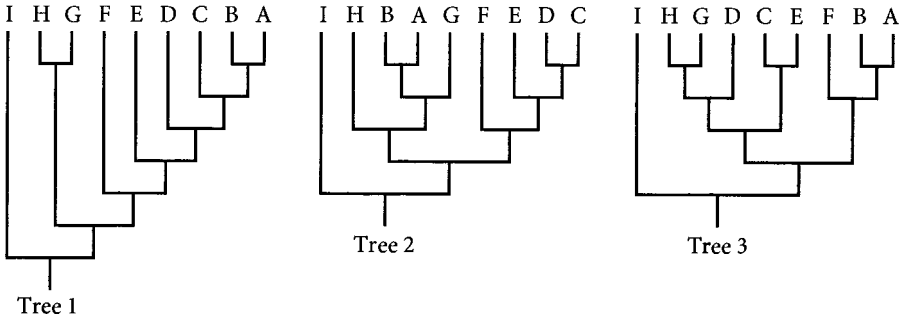


FIGURE 3.31 An example illustrating tree-to-tree distance. Tree 1 shares two clades with tree 2 and three clades with tree 3. It takes only one SPR rearrangement to convert tree 1 into tree 2 but three to convert it into tree 3.

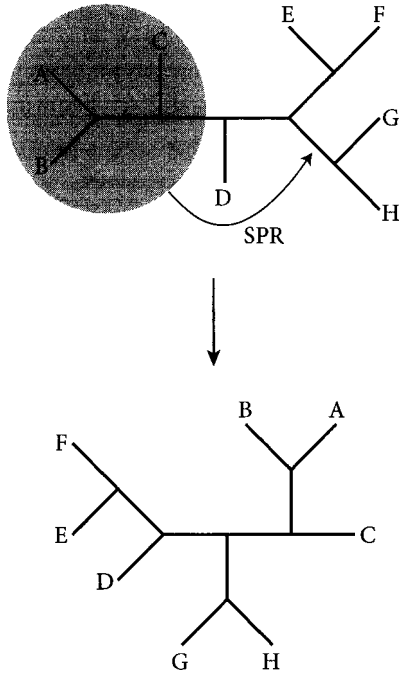


FIGURE 3.32 An example of subtree pruning and regrafting on an unrooted tree.

topology distance. In the case of Figure 3.31, converting tree 1 into tree 2 can be achieved by repositioning the (AB) clade. This shows that these two trees are separated by just one SPR, giving a distance of 1. Similarly, we can determine that it takes at least three SPRs to convert tree 1 into tree 3, giving those trees a distance of 3. Thus, in terms of SPR rearrangements, tree 1 is closer to tree 2 than to tree 3.

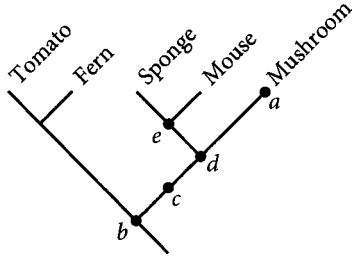
Before leaving tree-to-tree distances and tree rearrangement, it is worth noting that it is possible to convert one tree topology into *any* other tree topology by doing a series of SPR rearrangements. This tells us that there is a continuous, multidimensional "space" of tree topologies and that one can move through this space by rearranging trees SPR by SPR. As will be discussed more fully in Chapter 7, this ability to traverse tree space is what allows a computer program to search systematically for the optimal trees, even when the analyses include very many taxa.

FURTHER READING

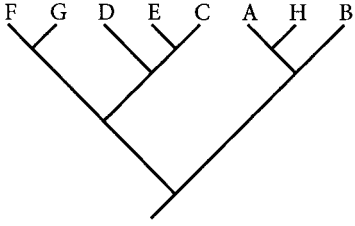
Interpretation of a phylogenetic tree: Maddison and Maddison 2000
Common problems in reading trees: Baum et al. 2005; Baum and Offner 2008; Gregory 2008; Catley and Novick 2008, 2009
Rooting trees: Maddison et al. 1984

CHAPTER 3 QUIZ

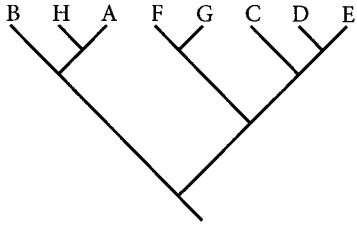
1. Which of the five labelled nodes in the tree corresponds to the most recent common ancestor of a mushroom and a sponge?



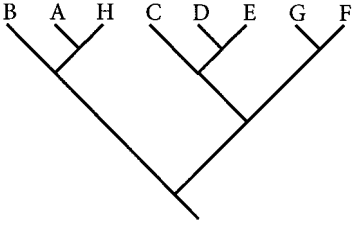
2. Which of the four trees depicts a different pattern of relationships than the others?



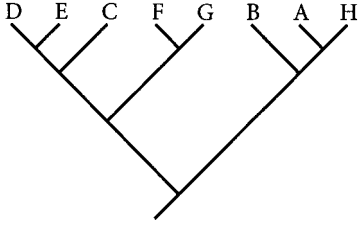
a



b

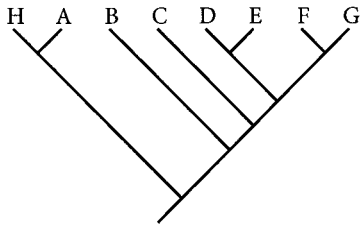


c

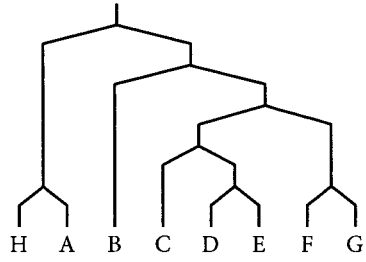


d

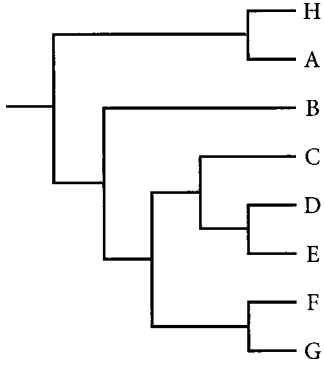
3. Which of the four trees depicts a different pattern of relationships than the others?



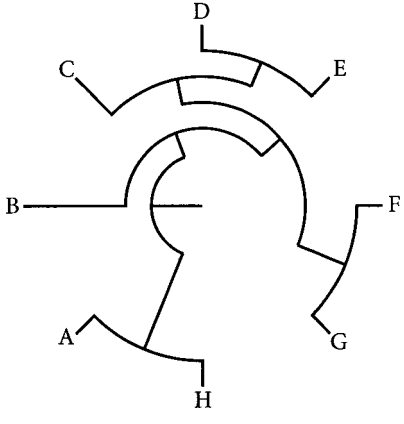
a



b

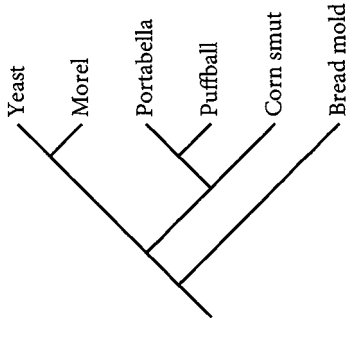


c



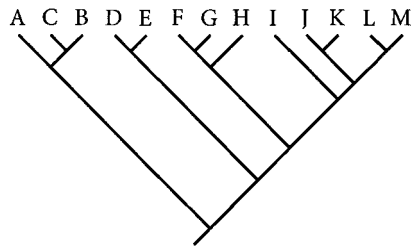
d

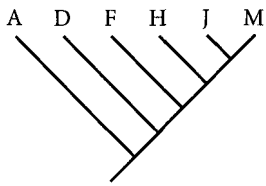
4. The clade with the name Dikaryomycota comprises all the descendants of the last common ancestor of morel and puffball. Which taxa on this tree are *not* in Dikaryomycota?
   a. Bread mold b. Corn smut
   c. Yeast, bread mold
   d. Yeast, portabella, corn smut, bread mold
   e. Yeast, corn smut, puffball



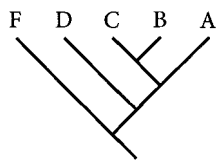
5. On the tree, which of the following sets do *not* form a clade?
   a. Yeast, morel b. Portabella, puffball c. Portabella, puffball, corn smut
   d. Portabella, puffball, corn smut, bread mold e. All taxa except bread mold

6. Which of the five smaller trees is compatible with the larger tree once the extra taxa are pruned?

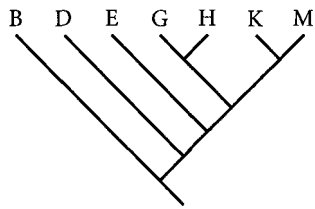




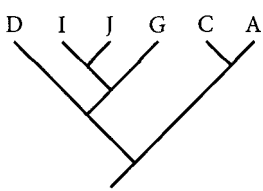
a



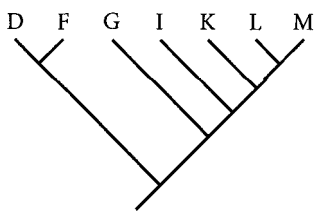
b



c



d

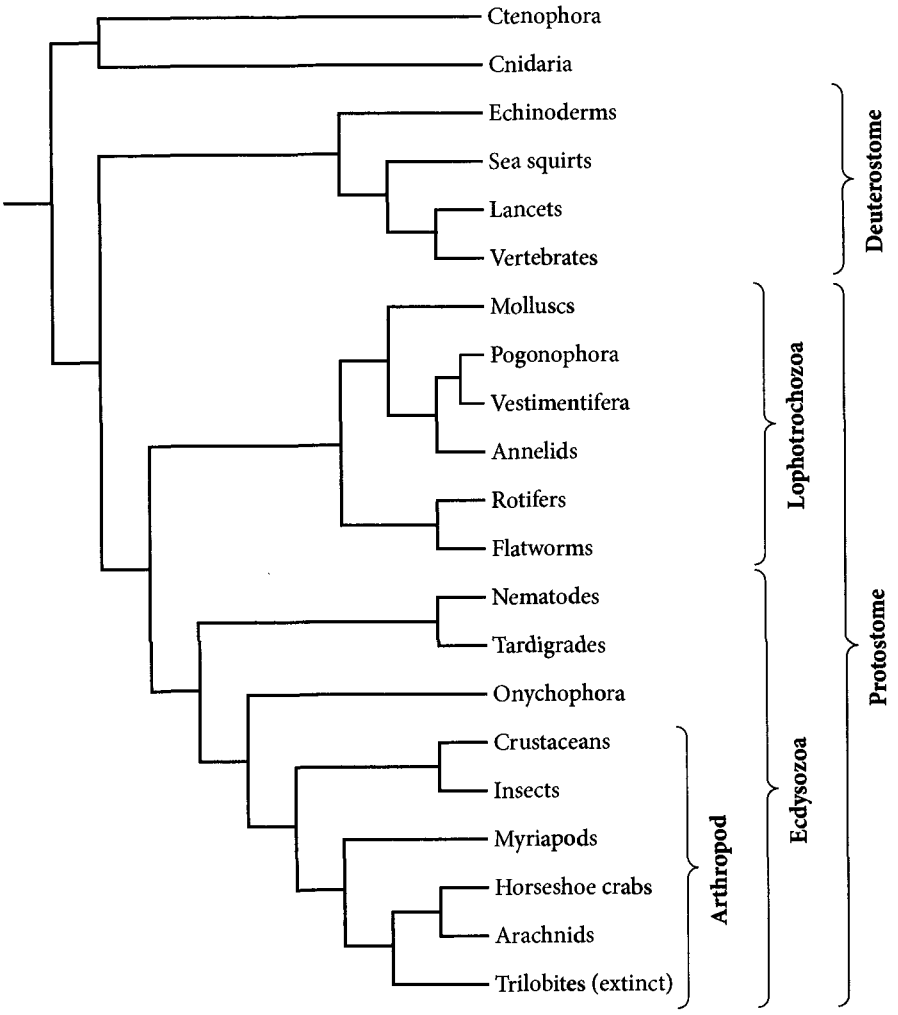


e

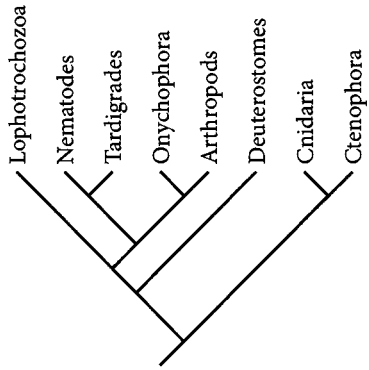
Questions 7–9. The tree on the facing page shows some hypothesized relationships among the major animal groups. Some major clades are named. You do not need to know these organisms to answer the questions.

7. Which of the following organisms are in Ecdysozoa but are not arthropods?
   a. Molluscs b. Flatworms c. Onychophora d. Crustaceans e. Arachnids

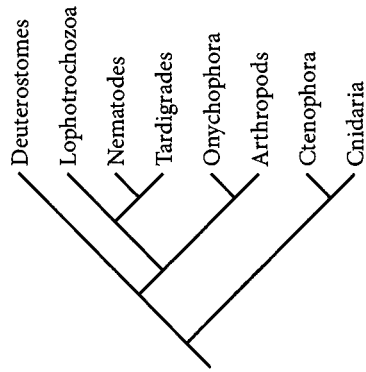
8. Each node on this tree is ancestral to somewhere between two and twenty-one of the tips. Which of the following nodes does not exist on this tree?
   a. Ancestral to annelids but not flatworms
   b. Ancestral to trilobites and ctenophora
   c. Ancestral to rotifers and echinoderms, but not cnidaria
   d. Ancestral to sea squirts and lancets, but not molluscs
   e. Ancestral to crustaceans and tardigrades, but not myriapods


Ctenophora
Cnidaria
Echinoderms
Sea squirts
Lancets
Vertebrates
Deuterostome
Molluscs
Pogonophora
Vestimentifera
Annelids
Rotifers
Flatworms
Lophotrochozoa
Nematodes
Tardigrades
Onychophora
Crustaceans
Insects
Myriapods
Horseshoe crabs
Arachnids
Trilobites (extinct)
Arthropod
Ecdysozoa
Protostome

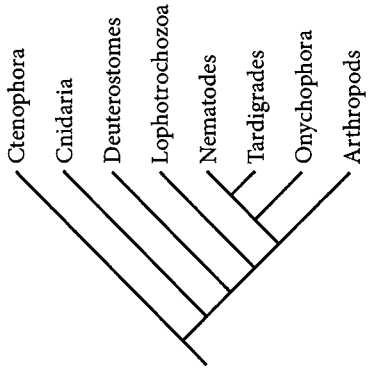
9. Supposing you merge each of the arthropod, Lophotrochozoa, and deuterostome clades into a single tip. Which of these trees would result?



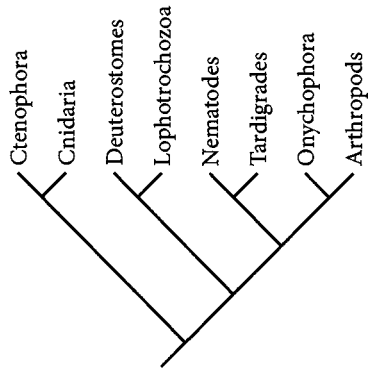
a



b

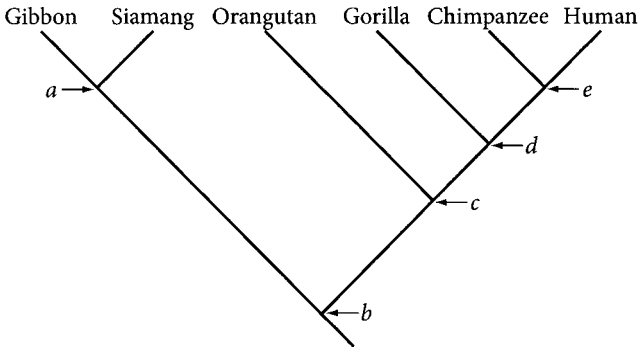


c



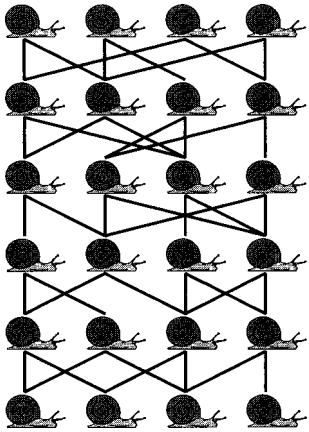
d

10. Given the tree shown on the next page, which (if any) of the following living species are ancestors of humans?
    a. Chimpanzee
    b. Orangutan
    c. White-handed gibbon
    d. All of the above
    e. None of the above (no living species)



11. On the same tree, which of the following is not necessarily true (cannot be assumed from the information given)?
    a. *a* is a descendant of *b*
    b. *d* is a descendant of *b*
    c. *e* lived after *b*
    d. *a* lived after *c*
    e. *c* lived before *d*

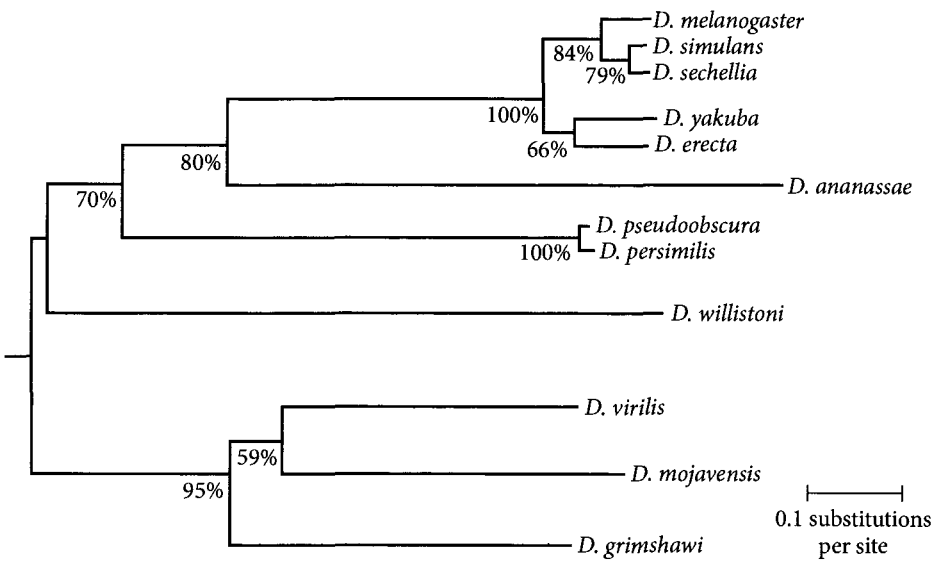
12. This figure shows a population lineage with six generations of snails (which, like most snails, are hermaphrodites). Which way is the time axis pointing in this lineage and how do you know?



13. Show your ability to manipulate trees.
    a. Draw an 8-taxon tree in a diagonal format.
    b. Redraw the same topology in rectangular format, with some branches rotated so that the taxon names are in a different order.
    c. Draw two different 5-taxon trees that are each compatible with the 8-taxon tree.

14. Draw the tree that is provided here in parenthetical notation:
(1,((2,(3,4)),(((5,(7,8)),6),(9,((10,11),12))))).

Questions 15–19. The phylogram is based on a tree by the Drosophila 12 genomes consortium (2007) and depicts the relationships inferred for twelve species of fruit fly (*Drosophila*) whose genomes were completely sequenced. Assume that the tree is correctly rooted.



15. The branches are drawn proportional to the number of mutations estimated to have happened along that branch. Assuming the tree and branch lengths are accurate, which species has accumulated the fewest evolutionary changes since the last common ancestor of all twelve species?

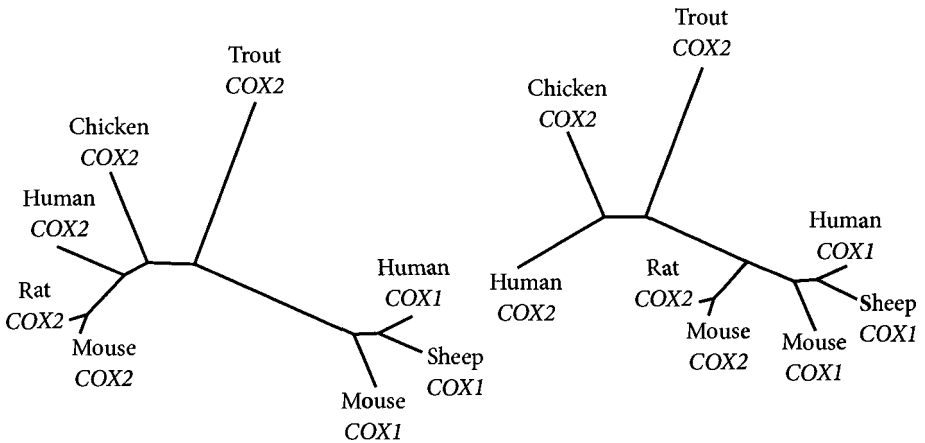
16. Which species has accumulated the most evolutionary changes?

17. The tree is annotated with bootstrap values that are over 50%. Which tips constitute a clade with an 84% bootstrap?

18. Of the annotated clades on this tree, which is the weakest (most uncertain)?

19. Suppose you decided that internal branches with bootstraps of less than or equal to 79% are not reliably resolved and you wanted to represent them as polytomies. Draw the resulting tree.

20. How many SPR events separate these two trees, and which taxa are involved?



CHAPTER FOUR

Trait Evolution

If the realm of phylogenetics were limited to descent from common ancestry, it would not be able to explain something as basic as the fact that living organisms are not all identical. Because evolution is about descent *with modification*, our conceptual framework needs to accommodate the phenomenon of trait evolution. ***Traits*** (also called characters, characteristics, or phenotypes) are features of organisms that arise through the expression of an organism's genetic makeup in a particular environment. In this chapter, we overlay trait evolution onto the concept of phylogenies offered in Chapter 3. We then discuss how changes in traits can be represented on tree diagrams and provide a brief introduction to the concept of homology.

TRAIT EVOLUTION IN A SINGLE LINEAGE

Imagine a population of sexually reproducing plants with petals that are dark red. Figure 4.1 shows such a population. This population is very small, but the principles scale up easily. In generation two, a mutation occurs in a gene that is involved in pigment production in flowers. The mutant form of the gene (the mutant ***allele***) cannot produce the red pigment. Individuals with two copies of the mutant allele (homozygotes) have yellowish-cream petals, due to the accumulation of the biochemical precursors to the red pigment.

In the figure, each organism is represented by two circles corresponding to the two alleles at the flower pigment locus. Black circles represent the functional allele, which produces red pigments, and gray circles represent the inactive alleles that produce no red pigment (and thus yield cream flowers in the homozygous condition). We will refer to the functional, red allele as the ***ancestral*** allele, because it was present in all members of the ancestral population. We will refer to the inactive, cream allele as the ***derived*** allele to communicate

the fact that it was derived from the ancestral allele by mutation. Individuals whose flowers are cream colored are surrounded by an oval.

As shown in Figure 4.1, the frequency of the derived allele increases over the course of several generations. The genetic composition of the population has changed: it has evolved. The speed with which the frequency of the cream allele increases suggests that it is favored by natural selection. However, evolution does not require selection; such changes in allele frequency can also occur by chance, a phenomenon called ***genetic drift***.

In this example, the ancestral red allele is ultimately lost from the population. In population genetics terms, we would say that the population has become ***fixed*** for the derived allele. It is worth highlighting that the loss of the ancestral red allele, which occurs when the derived allele goes to fixation, is irreversible. If the entire population complex (comprising all populations linked by occasional gene flow) becomes fixed for the derived allele, the ancestral allele could recur only by a reverse mutation, which is highly improbable.

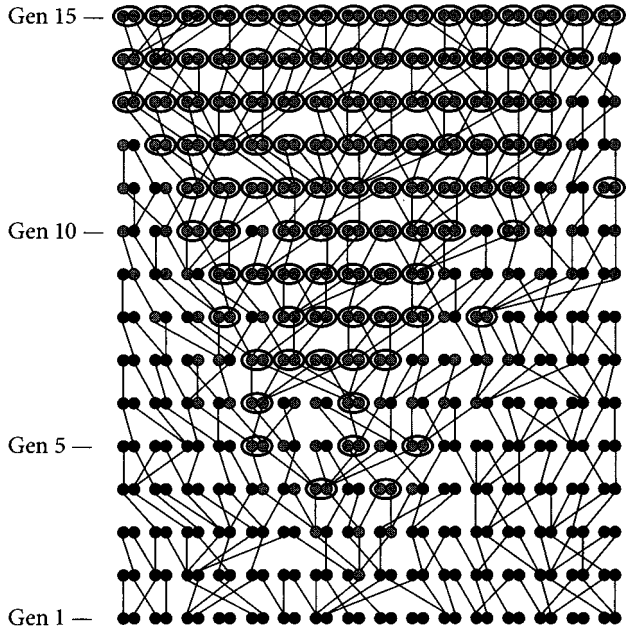


FIGURE 4.1 Fixation of a derived allele in a population over 15 generations.

Looking at the level of the organism's appearance over the generations depicted in Figure 4.1, phenotypic evolution has occurred. The population was originally fixed for the ***ancestral character state***, red petals. After 15 generations the population became fixed for the ***derived character state***, cream-colored petals. In short, we would say that cream-colored petals evolved from red petals.

The time from mutation to fixation of a derived character state will typically take much longer than 15 generations, especially when selection favoring the derived allele over the ancestral allele is weak (refer to population genetics or evolutionary biology textbooks for more details). Also, many derived character states require the accumulation of changes in multiple genes or multiple successive mutations at a single locus. It can take tens, hundreds, or thousands of generations to transition from a population fixed for an ancestral character state to one fixed for a derived state. Nonetheless, in the time frame of an entire phylogenetic tree, even 100,000 generations is a brief time interval. Therefore, it is often a convenient simplification when considering trees to view a derived trait as having arisen in an evolutionary instant. This is why trait evolution is most commonly depicted on phylogenetic trees as a line or bar across a branch. This does not mean that the derived character state arose and went to fixation in one generation. It simply indicates that trait evolution happened quickly relative to the rate of lineage splitting.

Many kinds of traits can evolve along phylogenetic lineages. All that is required is some degree of ***heritability***, which is defined as a tendency for offspring to resemble parents. It is probably obvious that the rules governing morphological (i.e., physical) traits, would apply also to heritable behavioral, physiological, and biochemical traits. In each case the trait is the result of a heritable genetic or developmental program unfolding in a particular environmental context. But one class of trait may require special clarification: geography.

Nearly all organisms have only limited dispersal ability, which means that geographic location is a heritable trait. A squirrel in Germany is relatively unlikely to have offspring that live in Ecuador, or even in Italy. Because of this dispersal constraint, geographic locale can often be marked on trees very much like more typical traits. For example, as discussed in Chapter 1, "living in the New World" probably evolved from "living in the Old World" along the branch leading to the New World monkey clade. As this case illustrates, biologists have made good use of the phylogenetic perspective to reconstruct the paths by which different species have acquired their geographic ranges (Chapter 11).

TRAIT EVOLUTION IN A BRANCHING TREE

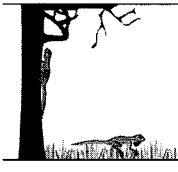
We noted above that once a mutant allele has gone to fixation, the ancestral allele is extremely unlikely to recur. This means that descendant lineages are all expected to possess the derived character state. This principle of descent with modification is all that is needed to extend our model of trait evolution from single lineages to whole branching phylogenies, whether of species or asexual clones. For example, once the ancestral mammal lineage became fixed for the production of hair, all its descendants inherited this trait. Although the color, length, texture, and distribution of hair subsequently evolved, the character state—"hair"—was retained through the many lineage-splitting events that gave rise to the approximately 5400 species of living mammals. Even dolphins and whales have some hair as fetuses.

In some cases a trait can evolve to resemble an earlier state, but this is not truly a reversal of evolution. Snakes have lost their limbs, but this does not mean that they have reverted to the state of the fishlike organisms that preceded the initial origin of walking legs: snakes do not have fins. And even in cases where a trait is indistinguishable from that of an ancestor, it is better viewed as a case of separate evolution rather than a reversal to the ancestral condition.

Before progressing further, let us clarify some terminology. In phylogenetics, it is conventional to distinguish ***characters*** from ***character states.*** A character is an attribute that potentially varies among the tips. For example, when considering mammals, hair color could be a character. Character states, in contrast, refer to alternative versions of the character that could occur in different organisms. For example, brown hair and black hair are alternative states of the character "hair color." It is important to note that characters and character states are understood relative to the variation seen among a set of taxa. If, for example, we were looking broadly across the land vertebrates, "hair" might be a character state of the character "integumentary outgrowth," whose alternative states might be "none," "scales," "feathers," or "hair."

To help visualize the process of trait evolution along the branches of a phylogenetic tree, consider a hypothetical example involving an ancestral lizard lineage, diversifying over time to yield six descendant species. Among the many characters that evolve somewhere on this phylogeny we track eight traits, whose ancestral character states are listed in Table 4.1. This example is based on Phylostrat, a computer program for exploring trait evolution along the branches of a phylogeny. The program is freely distributed by Jon Herron of the Univer-

TABLE 4.1 Characters and character states in lizards

	Character	Ancestral state	Derived state
	Crest on head	Absent	Present
	Colored collar	Absent	Present
	Preferred prey	Insects	Worms
	Pattern on back	Stripes	Mottled
	Tail spines	Absent	Present
	Habitat	Ground dwelling	Tree dwelling
	Tail spots	Present	Absent
	Dewlap (flap of skin under chin)	Absent	Present

sity of Washington (http://faculty.washington.edu/herronjc/SoftwareFolder/PhyloStrat.html).

Starting from one ancestral lizard lineage, five branching events occurred to give rise to six terminal species. In the course of this branching process, the traits evolved from the ancestral to various derived states on different lineages, as shown in Figure 4.2. Assume that this is the true tree and the actual history of trait evolution.

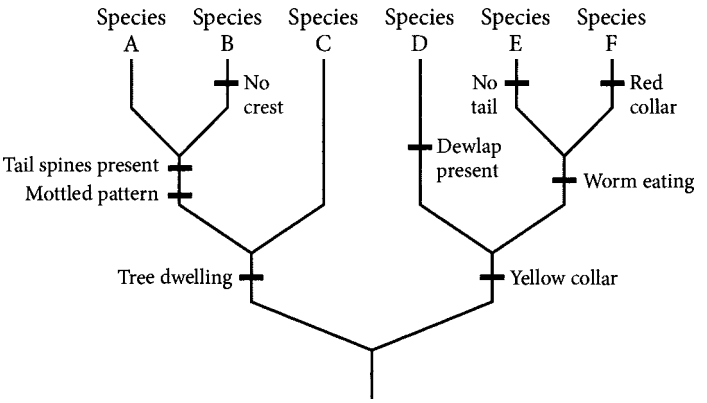


FIGURE 4.2 Trait evolution in six species of lizards. Table 4.1 lists the ancestral character states.

In the format we have chosen, character state changes are drawn with a mark at an arbitrary position along a branch. Changes marked on a lineage occurred somewhere along its length. Recall, from Chapter 3, that lineage branching primarily involves the geographic fragmentation of population lineages, which then *allows* the isolated lineages to undergo independent evolution. As a result, the evolution of derived traits that differentiate descendant lineages usually occurs after lineages have become genetically isolated. For this reason, we recommend against marking character state changes at nodes since such a placement might lead one to mistakenly assume that character evolution *caused* lineage splitting.

You should also notice that the distribution of changes is not entirely even; some branches have no changes, whereas one branch has two changes. Trait evolution is not a fully predictable phenomenon. Given enough time, some trait evolution is likely to occur in any descendant lineage, but it can occur at different rates in different lineages, and may or may not include changes in the particular subset of characters that we are considering.

In cases where multiple changes occur on the same lineage, we often do not know which happened first. The convention is to mark the two changes in an arbitrary order. For example, in Figure 4.2 the evolution of a mottled pattern is shown as having occurred before tail spines, but actually the order is probably unknown. When reading a tree that has multiple evolutionary changes marked

on a single branch, one should not assume that the order of events is accurately represented.

The figure conveys which character states occur in each of the six tips. Take the list of ancestral states, as found at the root, and then note the characters that changed on the path from the root to the tip. For example, individuals of species A have the ancestral states for five of the characters (head crest, insect eating, no collar, tail, and no dewlap) but have the derived states for three characters (living in trees, having a mottled pattern on their backs, and having spiny tails). You may find it useful to make up a list of the characteristics of the other five species and then check it against Table 4.2.

This example involves a few lizards, but the principles apply equally to the entire tree of life. The heritable features that we see in any living species, whether in morphology, physiology, biochemistry, geography, or behavior, are those traits that evolved somewhere in the species' ancestry. If you could trace a path from the origin of life through the full branching phylogeny to any tip, the set of traits seen in that tip would be the aggregate of all the derived traits evolving along that path. Organisms can be viewed as summations over their evolutionary history.

These considerations explain why trees are such important tools for organizing information about biological diversity. Since evolution occurs along a

TABLE 4.2 Character states in each of six lizard species (Sp. A–F) based on the character mapping in Figure 4.2

	Character states					
Character	**Sp. A**	**Sp. B**	**Sp. C**	**Sp. D**	**Sp. E**	**Sp. F**
Crest on head	Present	Absent	Present	Present	Present	Present
Colored collar	Absent	Absent	Absent	Yellow	Yellow	Red
Preferred prey	Insects	Insects	Insects	Insects	Worms	Worms
Pattern on back	Mottled	Mottled	Stripes	Stripes	Stripes	Stripes
Spiny tail	Present	Present	Absent	Absent	Absent	Absent
Habitat	Tree	Tree	Tree	Ground	Ground	Ground
Tail	Present	Present	Present	Present	Absent	Present
Dewlap	Absent	Absent	Absent	Present	Absent	Absent

tree, keeping track of the tree allows us to store information about diversity in a maximally efficient way. This can be seen quantitatively in this example. With eight traits and six species, you could memorize each species' state for each character, requiring you to store 48 pieces of data. Alternatively you could memorize the tree (composed of four clades), the eight ancestral states, plus the nine character state changes (listed beside the clade with which they are associated) for a total of 21 data points. You require less than half the brain space to store these same data using the tree framework than you would if you just memorized the data table.

As the number of tips and traits increases, the advantage of tree thinking becomes even more apparent. It is certainly easier to memorize that hair is a trait of all 5400 living species of the mammal clade than to have to put a mental check by the trait hair 5400 times. Furthermore, a single tree, for example, the tree of major vertebrate groups, can help you make sense of dozens of important traits.

Once you have developed an ability to think clearly about trees, you can start building a mental sketch of the full tree of life and gradually flesh it out by attaching more and more traits of interest. Any trait present in living organisms evolved somewhere on the tree of life. Traits found in all cellular organisms (the genetic code, glycolysis, DNA, ribosomes, etc.) can be attached to the ***stem lineage*** of all life (i.e., the lineage leading to the last common ancestor of all living organisms). Likewise the traits of particular living species will map at some depth within the tree, which tells you how widely that trait is shared with other organisms. Some human traits, such as written language or hunting with projectiles, evolved recently and are shared with no other living species. In contrast, other human traits, such as having a cell with a nucleus, map nearer to the root of the tree of life and are shared much more widely.

ANCESTRAL AND DERIVED CHARACTER STATES

While it is inappropriate to label an organism or taxon as either "ancestral" or "derived" (Chapter 3), this terminology is quite appropriate for character states. As discussed in relation to Figure 4.1, a derived character state is one that evolves from an ancestral character state due to fixation of derived alleles. Thus, any character with two alternative states within a certain group of organisms has a ***polarity***, a direction of evolution: one state is ancestral and the other

is derived. Determining a character's polarity is important from a phylogenetic perspective because ancestral and derived character states provide different kinds of phylogenetic evidence.

One way that scientists indicate that a concept is important is to associate it with new terminology. This is illustrated well by Willi Hennig, the father of phylogenetic systematics (Chapter 2), who coined novel terms for ancestral character states, ***plesiomorphies***, and derived character states, ***apomorphies***. Why, you may wonder, did Hennig think that these concepts deserved their own scientific-sounding terms?

When you find that a taxon is divided into two subgroups, one with character state 1 and the other with character state 2, your conclusions about phylogenetic history depend upon whether 1 or 2 is the apomorphic (derived) character state. As Hennig emphasized, shared, derived character states (***synapomorphies***) should be associated with clades. In contrast, groups of organisms sharing ancestral character states (***symplesiomorphies***) need not be clades. Figure 4.3 is reminiscent of Figure 3.10 but adds trait evolution to the picture.

In Figure 4.3a crocodiles and birds share a derived (apomorphic) character state, marked with a black bar. For example, the trait of laying eggs in nests that are tended by the parents may be an apomorphy shared by birds and crocodiles. This indicates that these taxa both descended from the branch on which the derived state arose, consistent with them forming a clade. In Figure 4.3b, lizards and crocodiles share a plesiomorphic or ancestral state (e.g., scales), while birds have an apomorphic or derived state (e.g., feathers). Although lizards and crocodiles have identical character states, because these traits are plesiomorphic, we cannot assume that lizards and crocodiles form a clade. Indeed, we know that lizards plus crocodiles is a non-monophyletic group because of

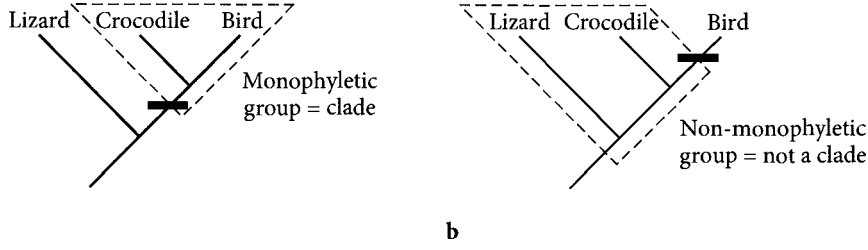


FIGURE 4.3 The association between character polarity and monophyly. Black bars indicate changes in character state.

other information, including traits whose evolutionary history resembles Figure 4.3a.

Using this logic, Hennig and many subsequent workers placed great emphasis on identifying synapomorphies. As described in Chapter 7, polarizing character states formed a critical step in Hennigian phylogeny reconstruction. While this method for phylogenetic inference is no longer used, it is still important to keep track of character polarity because only synapomorphic character states are expected to be associated with monophyletic groups.

THE EVOLUTION OF DNA SEQUENCES

The evolution of visible (phenotypic) characters is the result of changes at the molecular level—specifically, changes in the DNA sequence of some part of the genome. While not all evolutionary changes in a DNA sequence will result in phenotypic changes (some are ***silent***), almost all heritable changes in phenotypes are attributable to changes in DNA sequences. If phylogenetic trees are to provide a comprehensive framework for thinking about evolution, they must also accommodate molecular sequence evolution.

A DNA strand contains a series of nucleotides that is copied during DNA replication. Although genomic DNA occurs as a double-stranded helix, molecular evolution is typically modeled as a single strand. Because of complementary base pairing, the strand not shown can be determined entirely from the sequence of its complement. Figure 4.4 depicts one parental DNA strand of 13 bases, its two children, and its four grandchildren.

Each nucleotide position in a daughter sequence is copied from a particular position in its parent sequence. A nucleotide position in an offspring is defined to be ***homologous*** to that in a parent if the former was copied from the latter during DNA replication. Molecular homology of nucleotide positions is independent of the fact that copying is imperfect. A mutation occurring at a nucleotide position does not change that position's homology. For example, position number 7 is homologous in all the grandchildren despite the fact that two have A's and two have T's at this position. Pairing during DNA replication and the *tendency* for the daughter strand to match the parent strand is what defines molecular homology.

Sequence evolution involves changes in the nucleotides occupying particular positions in the sequence. As the shaded bases highlight, mutations have arisen

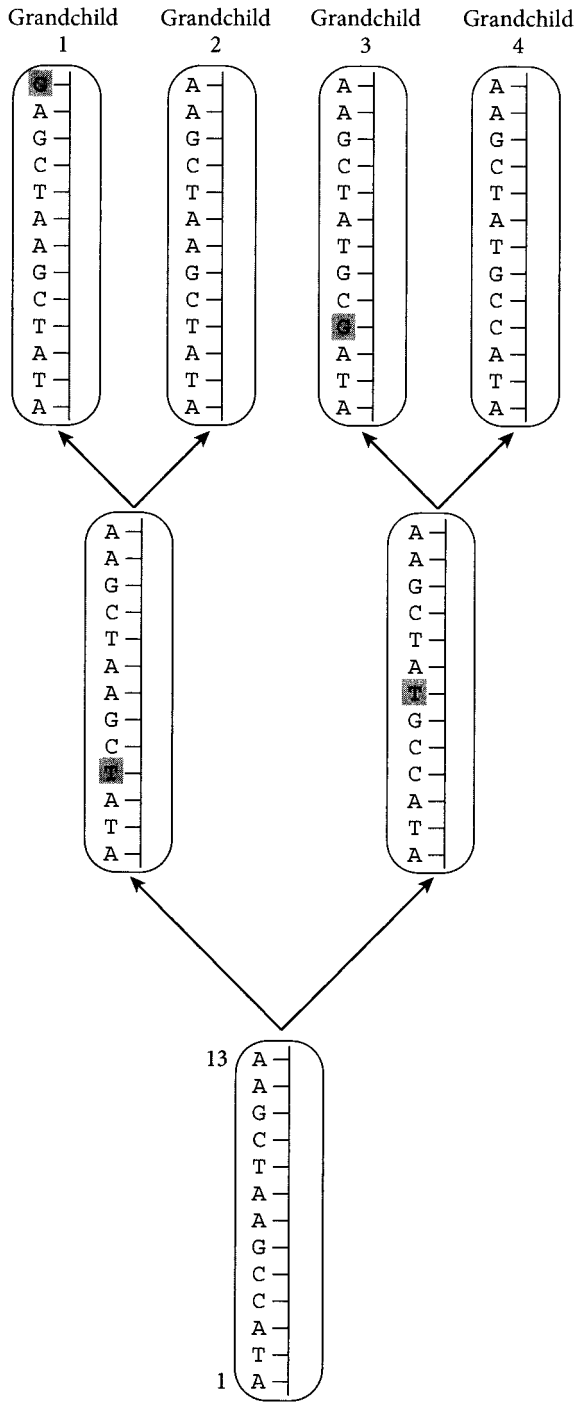


FIGURE 4.4 Changes in DNA sequences during two rounds of replication. Mutated positions are highlighted.

(either during DNA replication or at another time in the life of the organisms) so that the four grandchildren have distinct sequences from one another and from the ancestral sequence. We call the change of one base into a different base at the same position a ***substitution***. For illustrative purposes, we have assumed a higher rate of substitution than is typically seen in real data.

If you were given the ancestral sequence and told which changes had occurred on the lineage leading to grandchild 1, you could infer the sequence of grandchild 1. The principles here are exactly the same as for phenotypic traits such as tails and crests. That is to say, a sequence is the sum of all changes that occurred at some point during its evolution, overlaid upon the starting sequence.

The phylogenetic treatment of DNA sequences is very similar to the treatment of traits such as morphology. In the case of sequence evolution, each position in the sequence is a character and the bases occupying that position are character states. Thus, in reference to Figure 4.4, we could state that for character 13, grandchild 1 has the derived character state, G, whereas the other three grandchildren have the ancestral character state, A.

This description of molecular sequence evolution is somewhat simplified. We have ignored the fact that, as well as substitutions, previously exist-

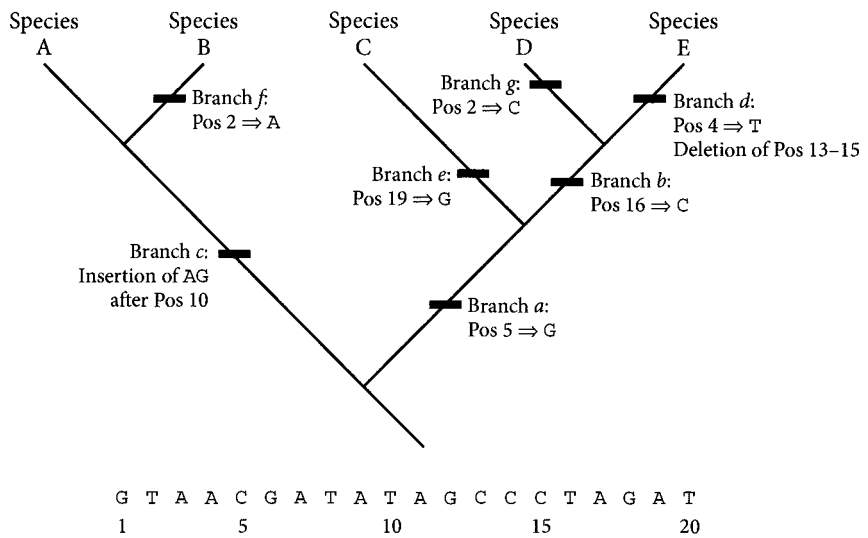


FIGURE 4.5 Example of DNA sequence evolution in five species. The ancestral sequence is provided at the bottom and subsequent changes are marked with black bars. Pos = Position.

ing bases can be deleted or new bases can be inserted into a sequence. When bases are inserted into a sequence, they have no locally homologous positions (though they may be copied from some distant part of the genome). Thus, from the point of view of this evolving sequence, these characters appeared from nowhere. An example of a ***deletion*** is shown on branch *d* in Figure 4.5, whereas branch *c* includes an ***insertion***. If we knew the true history of insertions and deletions, we could align the sequences such that homologous positions are above one another. Chapter 7 provides more details on sequence alignment and the treatment of insertion/deletion (***indel***) events. The correctly aligned sequences from the tips of Figure 4.5 are shown in Table 4.3.

While there are many similarities in the way that DNA sequence characters and other kinds of characters evolve along a tree, there is one important difference. The characters that make up a DNA sequence are physically connected in a specific order. The order in which morphological characters are listed in Table 4.2 is arbitrary, whereas the order in which the sequence characters (positions) are listed in Table 4.3 is constrained by their order in the sequence.

The physical connectedness of the positions in a DNA sequence might lead you to wonder whether their evolution is truly independent. Even under the assumption that an entire gene sequence shares the same gene tree (Chapter 6), evolutionary changes from one character state to another at one position are generally independent of character state changes at other positions in the sequence. The physical connections between adjacent positions do not, in themselves, result in nonindependent evolution. This principle explains why different positions in a DNA sequence, like different morphological characters, can be viewed as providing independent pieces of evidence on the structure of the true tree (Chapters 7 and 8).

TABLE 4.3 The sequences of the five terminal species shown in Figure 4.5

	Position																					
Taxon	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**10a**	**10b**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
Sp. A	G	T	A	A	C	G	A	T	A	T	A	G	A	G	C	C	C	T	A	G	A	T
Sp. B	G	A	A	A	C	G	A	T	A	T	A	G	A	G	C	C	C	T	A	G	A	T
Sp. C	G	T	A	A	G	G	A	T	A	T	-	-	A	G	C	C	C	T	A	G	G	T
Sp. D	G	C	A	A	G	G	A	T	A	T	-	-	A	G	C	C	C	C	A	G	A	T
Sp. E	G	T	A	T	G	G	A	T	A	T	-	-	A	G	-	-	-	C	A	G	A	T

AN INTRODUCTION TO HOMOLOGY

It is natural to look at two humans and note that both have the "same" characters, for example, nostrils. While that character may take on alternative states (e.g., varying in size or shape), nostrils themselves are taken to be equivalent in some deep way. Likewise, you might entertain the possibility that human nostrils are the "same" as the blowholes of whales. Conversely, you probably appreciate that the wings of insects and birds are not the "same," despite bearing the same name—wing—and being used for flight.

The technical term used to refer to evolutionarily equivalent characters in different organisms is ***homology***. As we saw earlier, nucleotide positions in two DNA molecules are considered homologous if they trace back to the same nucleotide position in an ancestor. Phenotypic traits like nostrils are not directly inherited—kids' noses are not direct copies of their parents' noses. Nonetheless, so long as a phenotypic trait is heritable, we can conceive of traits as coming from ancestors. Consequently, homology may be defined for phenotypic traits similarly to nucleotide positions: two traits are evolutionarily equivalent, that is, homologous, if they trace to the same trait in a common ancestor.

Given this definition of homology, the claim that nostrils and blowholes are homologous would be supported by showing that the common ancestor of whales and humans had nostrils that became modified over evolutionary time to become the human and whale versions. Likewise, the lack of homology of bird and insect wings would be supported by evidence that the last common ancestor of birds and insects did not have wings. This evidence could come from phylogenetic analysis (as in the examples below) but also from analysis of the trait's structure and development. There are some complications, which make homology one of the most difficult concepts in evolutionary biology. However, the key point here is simply that, because homology invokes common ancestry, it is an inherently phylogenetic concept.

To clarify the relationship between phylogenetic trees and homology, we will consider an example involving nectar-containing spurs in some hypothetical flowering plant species. We will run through three possible histories of trait evolution and we will ask, In which cases is the spur of species D homologous to the spur in species H?

Figure 4.6 depicts a case in which the last common ancestor of species D and H (indicated by the circle) had a spur, and both lineages retained this spur

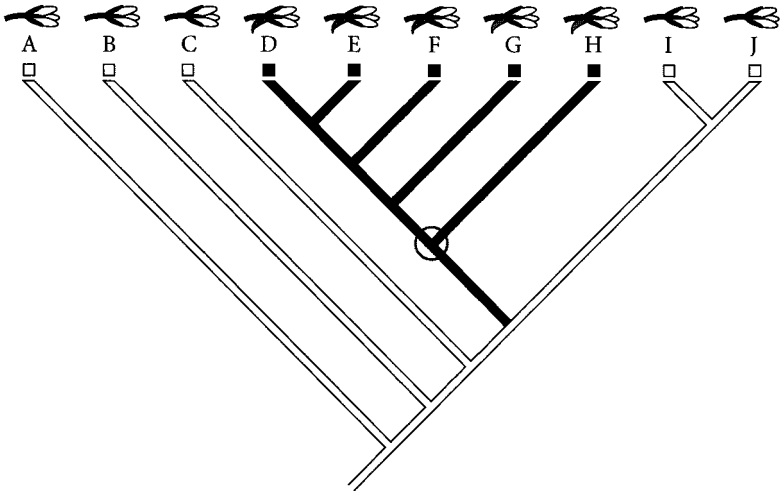


FIGURE 4.6 An evolutionary scenario under which the spurs in species D and H are homologous. Branches in black ended with spurred flowers, branches in white ended with spurless flowers.

from that point on. We have used a common format for depicting character evolution in which branches are colored based on their final character state (the state at the tip or just before the branch divided into two descendant lineages). Based on this history of trait evolution, the spurs in species D and H are homologous.

In Figure 4.7, the last common ancestor of D and H (again marked with a circle) lacked spurs. Spurs were acquired independently in the lineages leading to D and H. Because the spurs in D and H arose independently, they are the products of ***convergent evolution***. As a result, they are not homologous. This would be true even in the unlikely event that these independently evolved spurs utilized a similar developmental pathway to achieve their distinctive form. While similarity of development and structure provide good initial evidence that traits are homologous, homology is defined based on common ancestry, not similarity.

Figure 4.8 shows a case in which the common ancestor of D and H had a spur but, nonetheless, the traits are not homologous. This is because, on the lineage leading to D, the original spur was lost, but then a similar structure

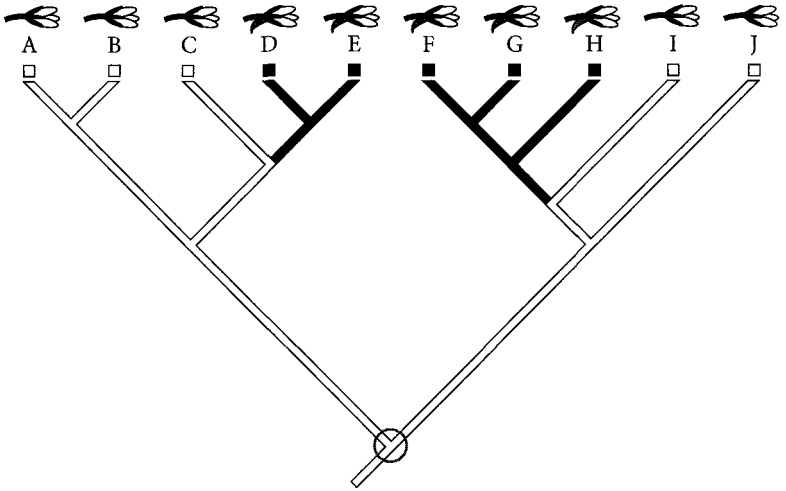


FIGURE 4.7 An evolutionary scenario under which the spurs in species D and H are not homologous due to convergence. Branches in black ended with spurred flowers, branches in white ended with spurless flowers.

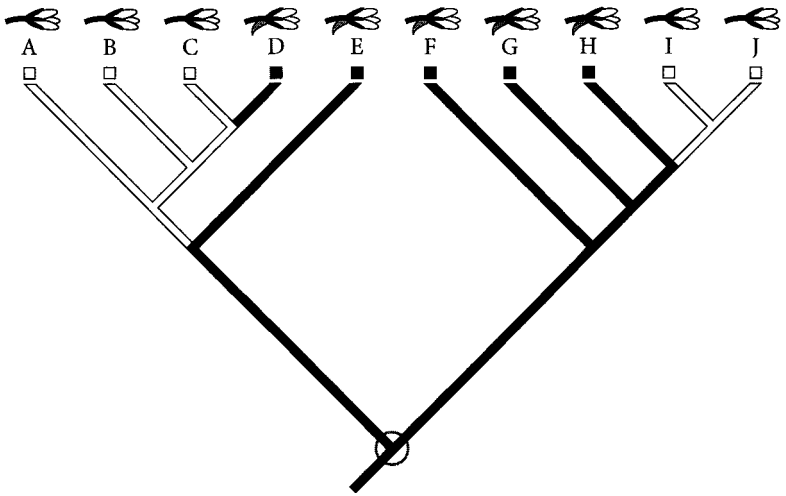


FIGURE 4.8 An evolutionary scenario under which the spurs in species D and H are not homologous due to reversal. Branches in black ended with spurred flowers, branches in white ended with spurless flowers.

evolved again. This pattern is sometimes called a ***reversal***. Because the spurs in D and H trace to two different evolutionary origins, they are not homologous.

HOMOPLASY AND CONSISTENCY

Homoplasy applies whenever character states are found to arise more than once on a given tree. Convergence, as shown in Figure 4.7, is a kind of homoplasy in which a derived trait arises multiple times on the same tree. Similarly, reversal, as shown in Figure 4.8, is a kind of homoplasy.

Homoplasy is sometimes presented as the opposite of homology, although this is inaccurate. Homology refers to a relationship between traits in different taxa, whereas homoplasy refers to the relationship between trait variation and a specific tree. While spur evolution in Figures 4.7 and 4.8 is homoplastic, in both cases the spurs in species F and G are homologous.

The opposite of homoplasy is not homology but ***consistency***. A character is said to be ***consistent*** with a tree if it evolved on that tree without any extra changes of character state. A character with two states (e.g., spur present vs. spur absent) is consistent with any tree that can explain the trait's evolution with one change of state.

Spur presence/absence is consistent with the tree in Figure 4.6 because there is only one change in character state. This logic can be extended to traits with more character states. In this case, consistency holds if the number of trait changes is one less than the number of character states. This ensures that each derived character state arose just once. If the number of character state changes is equal to or greater than the number of character states, then the character shows some homoplasy.

The idea of consistency is illustrated with some DNA characters in Table 4.4, all of which evolved along the tree shown in Figure 4.9. In the characters on the left, the number of character state changes on the tree is exactly one less than the number of character states observed. Therefore, these characters are consistent with the tree. In contrast, the characters on the right experienced as many or more character state changes than character states, meaning that these characters show homoplasy.

A useful way to quantify the consistency of a character with a tree is by using the consistency index, or CI. This index reports the minimum number of changes needed to explain a trait's evolution (L_{min}), which is one less than

the number of character states, divided by the actual number of character state changes (L_{obs}). This is summarized in the equation $CI = L_{min} / L_{obs}$. For example, if the minimum number of changes to explain a character is 3 but the observed number is 4, then CI = ¾ = 0.75. Table 4.4 provides the CI for each trait.

The CI is not the only measure of fit between character state variation and trees. You may encounter the homoplasy index, HI, which is equal to one minus the CI: HI = 1 − CI. The retention index, RI, is also widely used. This is a more complex index that corrects for the maximum number of steps that a character could have on any tree (L_{max}). Specifically, $RI = (L_{max} - L_{obs})/(L_{max} - L_{min})$. The

TABLE 4.4 Examples of DNA sequence characters that are either consistent or homoplastic (= homoplasious), given a particular tree (Figure 4.9).

Taxa	Consistent					Homoplastic				
A	A	G	T	G	G	G	C	G	A	T
B	A	G	T	G	G	G	T	T	C	G
C	A	G	T	G	C	A	T	G	G	A
D	A	G	T	T	T	A	C	T	C	T
E	A	A	C	T	A	G	C	A	A	C
F	A	G	C	C	A	G	T	A	G	G
States	1	2	2	3	4	2	2	3	3	4
Changes	0	1	1	2	3	2	3	3	4	4
CI	1.0	1.0	1.0	1.0	1.0	0.5	0.33	0.66	0.5	0.75

FIGURE 4.9 **The tree assessed for consistency in Table 4.4.**

RI is sometimes favored because, unlike the CI, it uses the full range from 0 to 1 (the minimum value of the CI varies, but is always well above 0.0).

The consistency index and related indices provide convenient measures of the fit between a character and a tree. The higher the CI or RI, the less homoplasy is implied. One application, introduced in Chapter 2, is to see if the amount of treelike structure in a complete data set is more than would be expected if the characters had not evolved up a tree. This is usually done by first finding the tree with the highest average CI for a data set (the ***most parsimonious tree***; see Chapter 7). Next, this CI is compared to the CI we would expect under the scenario that the data lack any treelike structure. Chapter 9 introduces one way to conduct this analysis, the so-called permutation tail probability (PTP) test.

PARSIMONY AS A WAY TO INFER THE HISTORY OF TRAITS

So far we have discussed trait evolution as it happens in theory. We have described how genetic changes in populations yield phenotypic variation and how phenotypes evolve over time such that some traits come to characterize clades of the tree of life. However, even when we are confident in the tree, we do not have direct knowledge of the evolutionary history of traits. Rather, when we encounter statements about the evolutionary history of a character, we are dealing with ***inferences***—informed conclusions guided by observational evidence.

In this section we provide a brief introduction to the use of the maximum parsimony criterion for making inferences about trait evolution. While there are more sophisticated methods now available (Chapter 10), parsimony is intuitive and provides an easy-to-understand introduction to the principles by which trait evolution can be reconstructed given some information about phylogenetic history.

As an example, let us look at the trait "wings" in insects. Figure 4.10 depicts the likely relationship among the winged insects and their closest relatives. Wings occur in adult damselflies, dragonflies, mayflies, and in the Neoptera, the large clade that includes such familiar insects as beetles, flies, bees, crickets, and butterflies. While a few neopteran lineages, such as fleas and lice, later lost wings, the ancestral condition at the base of Neoptera can be assumed to be wings present. Given this tree and the presence of wings in damselflies, dragonflies, mayflies, and Neoptera, what is the evolutionary history of wings?

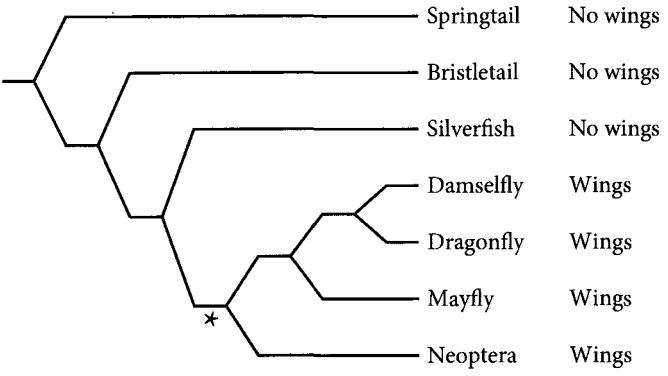


FIGURE 4.10 Phylogeny of Neoptera and related insects. Presence/absence of wings is indicated for each lineage. Based on parsimony, the most plausible hypothesis is that wings evolved once, on the branch marked with an asterisk.

There are many possible histories that could explain the distribution of wings. There could have been an ancient origin of wings at the base of the tree, followed by losses in springtails, bristletails, and silverfish; or there could have been four origins of wings, once each in dragonflies, damselflies, mayflies, and other insects; or there could have been a single origin of wings just before the last common ancestor of dragonflies, damselflies, mayflies, and Neoptera; and so on. All of these are valid explanations in that each tip has a state that is a summation over its evolutionary history. But only one of these alternatives actually happened. How can we decide which explanation is the best working hypothesis?

A simple way to pick the most plausible history is to apply the ***principle of parsimony*** (we will talk more about this concept in Chapter 7). The logic of parsimony may be familiar under the label Occam's (or Ockham's) razor, named after the fourteenth-century logician and friar William of Ockham (ca. 1288–1348), who said, "pluralitas non est ponenda sine necessitate" (plurality should not be posited without necessity). This principle holds that the simplest explanation of a phenomenon is the most likely to be true.

Suppose you are sitting at a police switchboard in a North American city and within a span of five minutes you receive two calls reporting a tiger on the loose. Theoretically this could indicate the existence of two tigers roaming the

streets. But would you really suspect this? No. More likely you would assume that there is just one tiger and two independent witnesses. The reason this is the logical conclusion is that calls reporting tigers are very rare. As a result, it is more likely that the two calls represent two manifestations of the same rare event rather than two independent, rare events.

We can use the same principle to make an inference about the evolution of wings. Gaining or losing wings during evolution is rare. Therefore, the most plausible explanation of the observed pattern is the one that requires the fewest evolutionary origins (or losses) of wings. Given the tree shown in Figure 4.10, the most parsimonious scenario is that wings evolved once on the branch marked with an asterisk. Thus, it is most parsimonious to assume that the wings of dragonflies, mayflies, damselflies, and Neoptera are homologous and that their sister group, silverfish, does not have any winged ancestors.

A single origin of wings is the most parsimonious explanation because it can explain the distribution of wings among the tips with only a single change in character state. This reconstruction is the one that maximizes the consistency index. In contrast all other scenarios require additional gains and/or losses of wings and are associated with more homoplasy and lower consistency indices.

Figure 4.11 considers four other scenarios for the evolution of wings. Gains of wings are marked with a plus sign (+) and losses of wings are marked with a minus sign (−). The cases to the right involve four changes of state (four gains or one gain and three losses), whereas the cases to the left involve two changes of state (two gains or one gain and one loss). Because these scenarios involve more than one change of character state, they are, by definition, less parsimonious than the hypothesis of a single origin of wings. Extending this logic, it is also fair to say that the two scenarios on the right are less likely than the two on the left because they require the occurrence of even more rare events (and have even lower CIs).

In the case of wing evolution in insects, there is a single reconstruction of character history that is more parsimonious than any other. However, sometimes two or more alternative reconstructions of character history are equally parsimonious: each involves the same number of evolutionary events. For example, consider the evolution of echolocation, an ability to "see" by emitting sounds and listening for echoes, in bats.

Echolocation is well developed in the narrowly defined Microchiroptera ("microbats") as well as the horseshoe bats and their relatives, but is absent

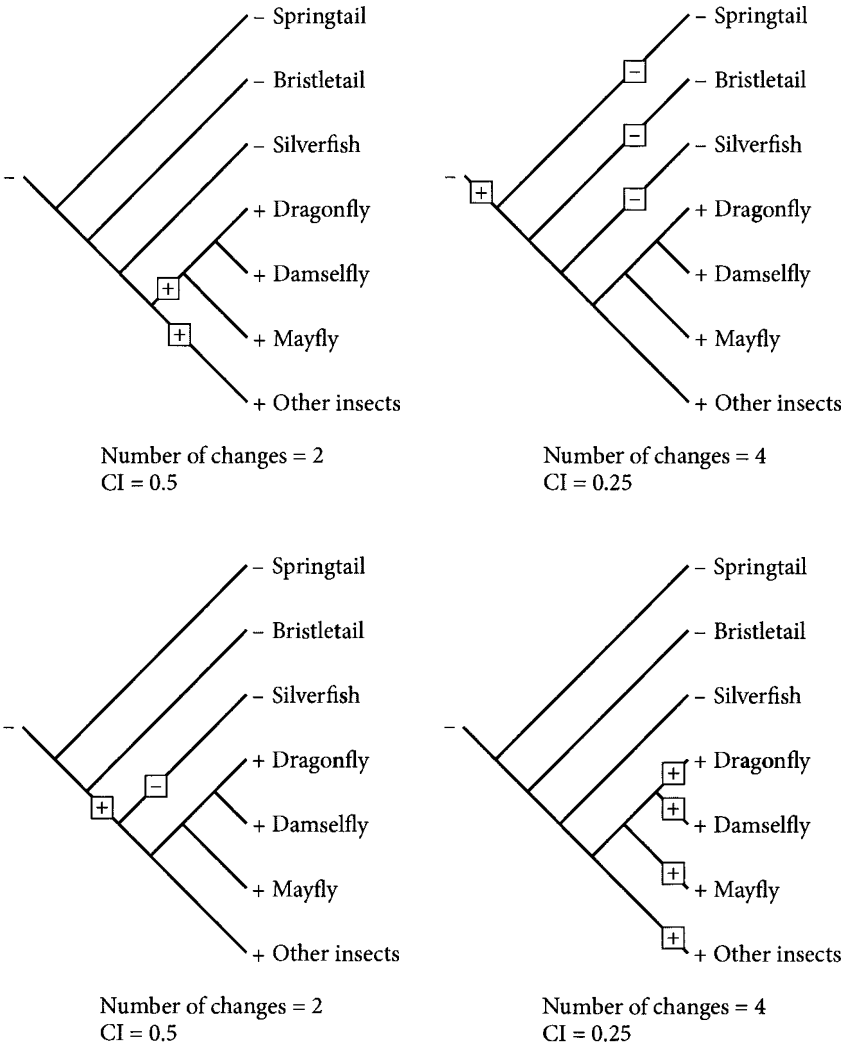


FIGURE 4.11 Four of the possible scenarios for the evolution of wings in insects. The presence of wings in the tips is indicated "+" and their absence by "–." Gains and losses of wings under each scenario are indicated using the same convention. The number of changes and the consistency index are indicated for each scenario.

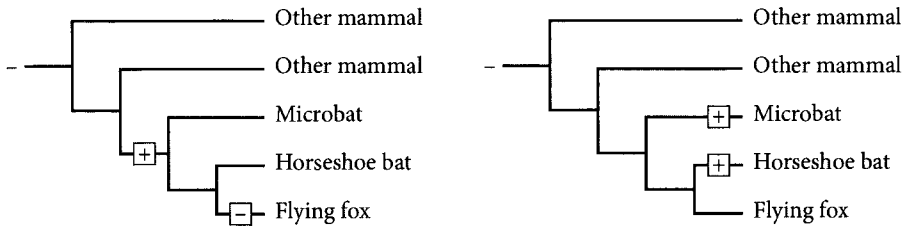


FIGURE 4.12 **Two equally parsimonious scenarios for the evolution of echolocation in bats.** Gains of echolocation are indicated with "+" and losses with "–."

from the flying fox (fruit bat) group. Given the phylogeny shown in Figure 4.12 (from Teeling et al. 2005), there is no way to explain the evolution of echolocation with fewer than two evolutionary state changes. But there are two equally parsimonious explanations. There could have been a single origin of echolocation close to the origin of bats, followed by a loss of this ability in the flying foxes. Or there could have been separate origins of echolocation in "microbats" and horseshoe bats. The principle of parsimony in its most basic form would not allow us to favor one or the other of these two explanations.

With further information about echolocation in different kinds of bats you might develop a preference for one hypothesis or the other. For example, if you concluded that independent gains of echolocation are highly improbable, whereas secondarily losing echolocation is more likely, you might favor the evolutionary scenario on the left. But if all you go on is simple parsimony, and if you assume that this tree is correct, both scenarios are equally plausible.

In this chapter we have explored the evolution of traits along the branches of phylogenetic trees. It should now be clear that by using parsimony (and more sophisticated methods to be introduced in Chapter 10) we can combine trees with knowledge of the distribution of character states among tips to learn about the history of trait evolution. Equally importantly, knowledge of trait evolution forms the basis of phylogenetic inference. By understanding the evolution of morphological and molecular characters, we are able to develop rigorous methods to reconstruct phylogenetic trees that are reasonable hypotheses of evolutionary history. We delve into these methods starting in Chapter 7.

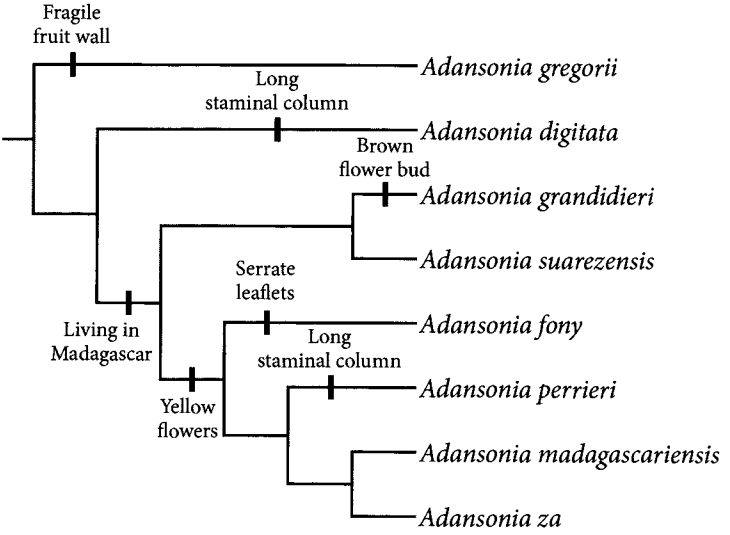
FURTHER READING

Trait evolution: Maddison and Maddison 2000
Molecular evolution: Page and Holmes 1998; Graur and Li 2000
Character polarity: Hennig 1966
Homology (conceptual issues): Patterson 1982; Wagner 1989; Abouheif 1997; Mindell and Meyer 2001; Hall 2003; Scotland 2010
Homoplasy and consistency: Farris 1989; Kitching et al. 1998; Wake et al. 2011
Parsimony: Sober 1983, 1991

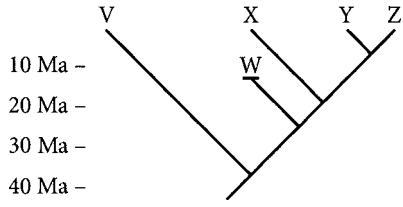
CHAPTER 4 QUIZ

1. On a phylogenetic tree, where are character state changes typically marked?
   a. at the root b. at nodes c. along branches d. at the tips
   e. character state changes are not shown on trees

2. Consider a plausible phylogeny of the baobab trees, genus *Adansonia*. On the basis of this tree, which species lives in Madagascar, has woody (not fragile) fruit walls, white (not yellow) flowers, and green (not brown) flower buds?
   a. *A. gregorii* b. *A. digitata* c. *A. grandidieri* d. *A. suarezensis*
   e. *A. fony*

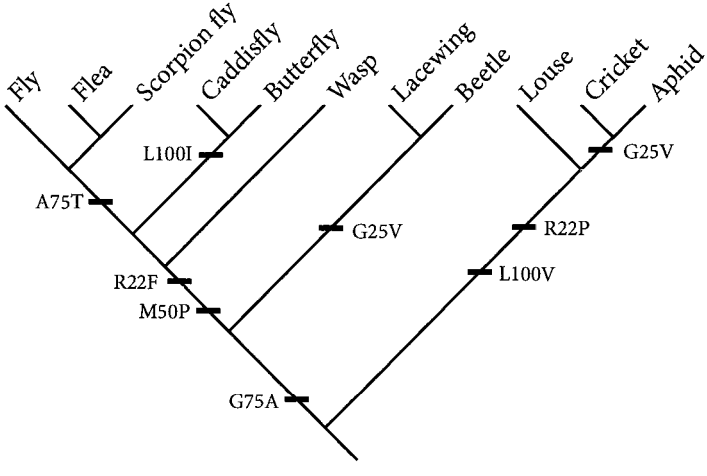


3. Which statement is best supported by the chronogram?

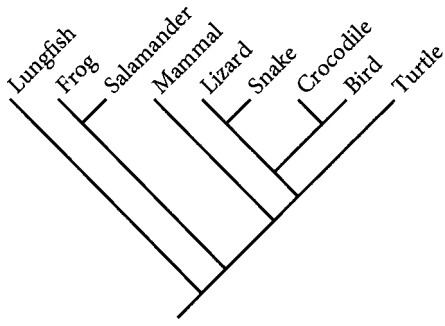


   a. Traits occurring in species W will be more ancient than those occurring in species Z
   b. Traits occurring in species Z will tend to be more advanced than those occurring in species Y
   c. Traits in species X will tend to be more similar to the traits of species W than to the traits of species Z
   d. Traits occurring in species W and X will tend also to occur in Z
   e. V has the most plesiomorphic character states

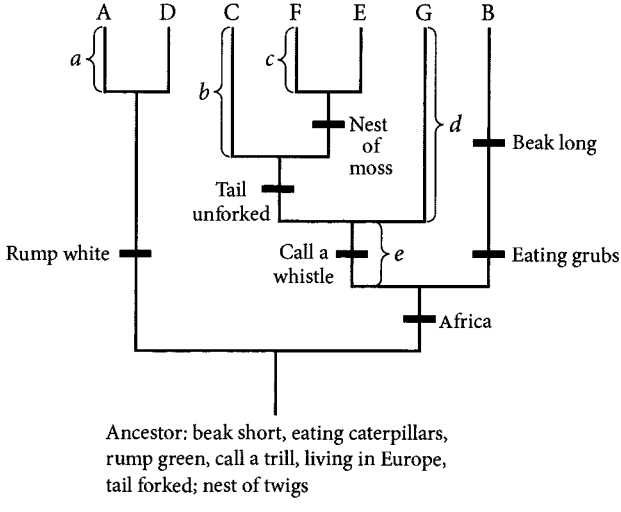
4. Consider a plausible phylogeny for some of the major insect groups. Along the branches are marked changes in a hypothetical protein sequence. The marks list the amino acid position in the protein that changed, flanked by the standard one-letter codes for the amino acids before the change (to the left of the number) and after the change (to the right of the number). For example, L122S means that position 122 changed from a leucine (L) to a serine (S). Based on this tree, what amino acids does a butterfly have at positions 50, 75, and 100?
   a. 50 = P (proline); 75 = A (alanine); 100 = I (isoleucine)
   b. 50 = M (methionine); 75 = G (glycine); 100 = L (leucine)
   c. 50 = R (arginine); 75 = G (glycine); 100 = V (valine)
   d. 50 = P (proline); 75 = T (threonine); 100 =V (valine)
   e. 50 = F (phenylalanine); 75 = T (threonine); 100 = P (proline)



5. Consider the following three facts: (1) the tree shown is correct; (2) The amnion (a membrane that surrounds the embryo) evolved once and was never lost; (3) A snake and a turtle have an amnion. Which of the following *must* also be correct?
   a. Salamanders have amnions
   b. Birds have amnions
   c. Mammals have amnions
   d. Lungfish do not have amnions
   e. Mammals do not have amnions



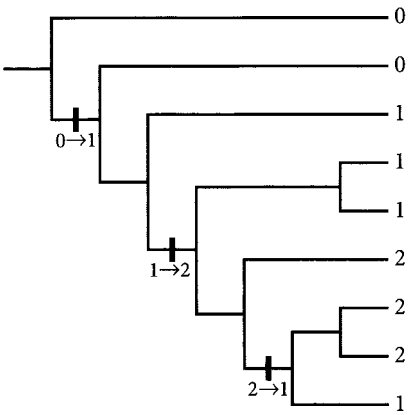
Questions 6–7 refer to the tree for a hypothetical group of birds.



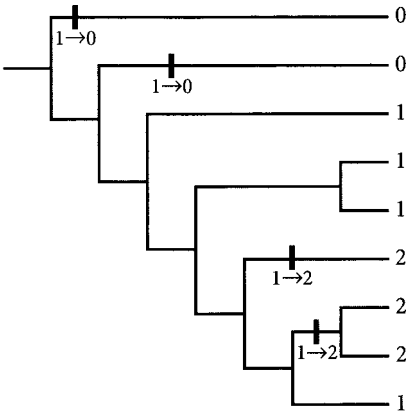
6. Which of the following sets of traits do you expect to find in bird E?
   a. Rump white, Africa, call a whistle, tail unforked, beak long
   b. Eating caterpillars, rump green, Europe, nest of moss
   c. Nest of moss, forked tail, rump white, Africa
   d. Beak short, eating caterpillars, rump green, tail unforked

7. Which of the labeled branches (*a*, *b*, *c*, *d*, or *e*) contain at least some individuals with the following combination of traits: eating caterpillars, rump green, Africa, and call a trill?

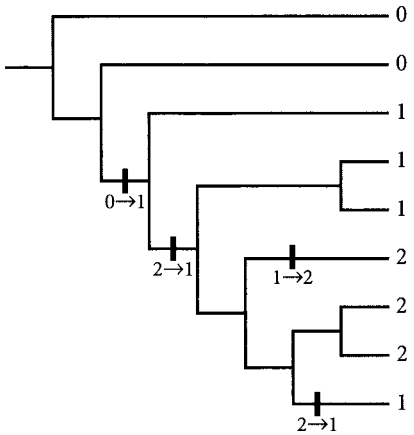
8. The four trees each depict a history of trait evolution for a character with three characters states, 0, 1, and 2. The character states at the tips are shown, as well as each transition. Which of these four provides a complete and plausible history of trait evolution?



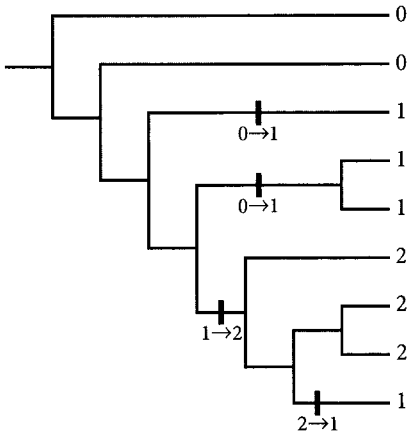
a



b

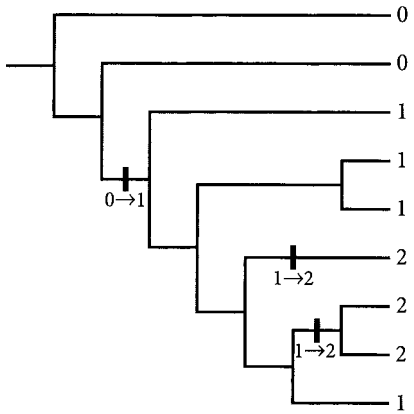


c

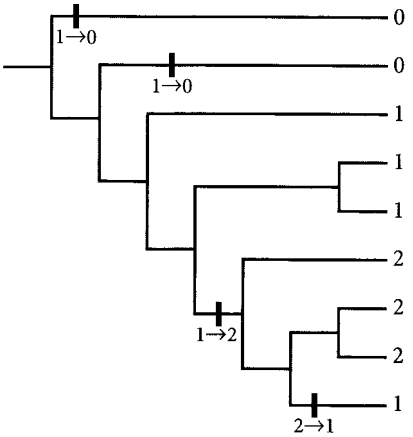


d

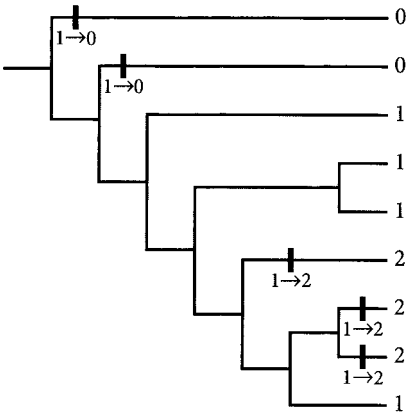
9. The four scenarios for trait evolution all validly explain the observed trait variation on this tree. Which is most parsimonious?



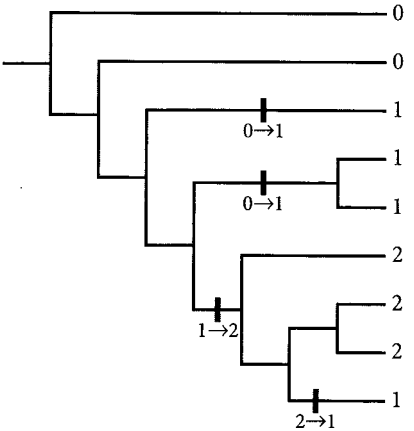
a



b

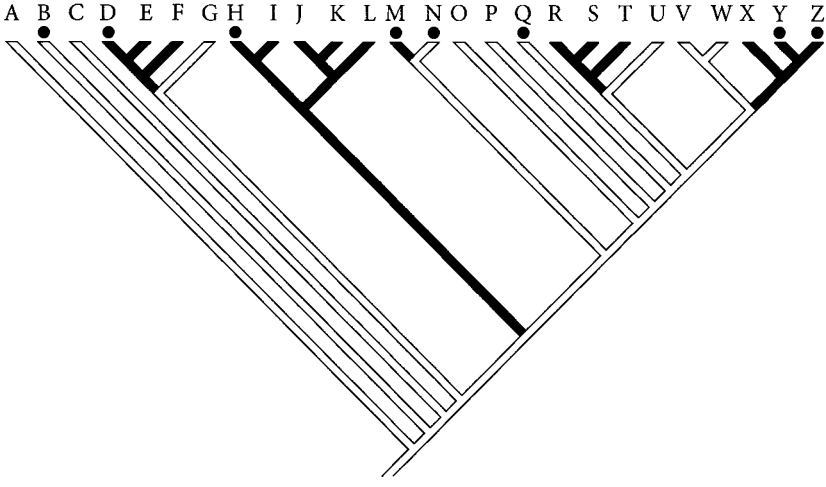


c



d

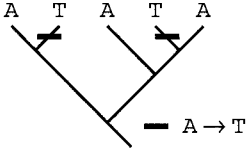
Questions 10–11. In the tree, white branches depict lineages that live in Africa, whereas the others live in South America. The taxa marked with a black circle have pink flowers, whereas the rest have yellow flowers.



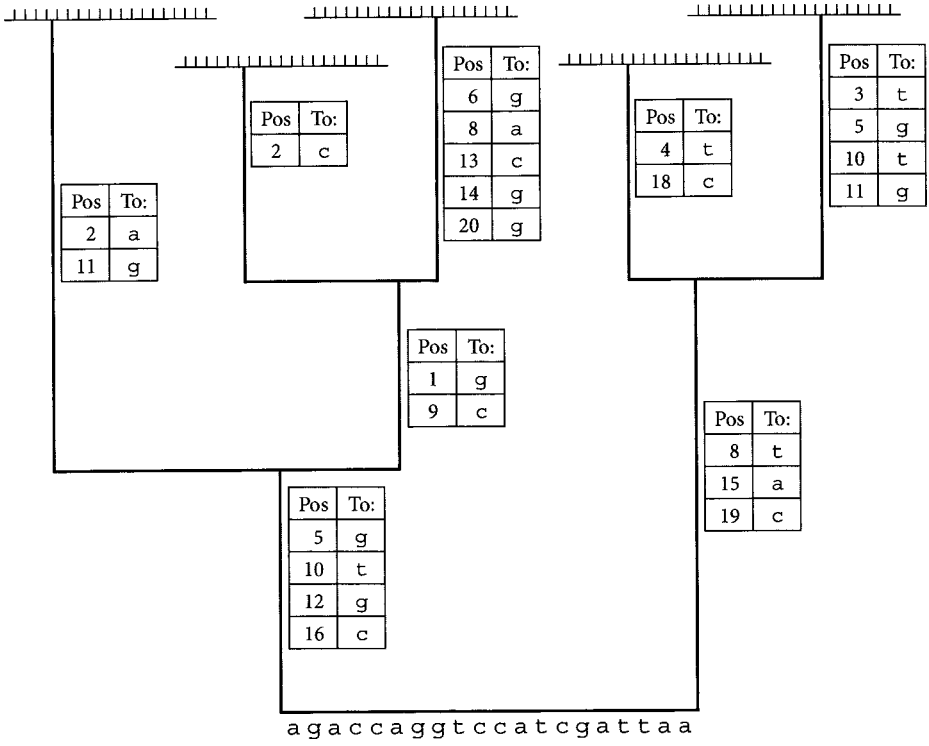
10. Using parsimony, how many times did pink flowers evolve?
    a. 1 b. 3 c. 4 d. 6 e. 8

11. How many of those origins of pink flowers occurred in populations that were living in Africa?
    a. 0 b. 1 c. 3 d. 4 e. 6

12. This tree shows the state of each tip for a particular position in a DNA molecule. What is the consistency index for this position?
    a. 0.2 b. 0.25 c. 0.4 d. 0.5 e. 2



13. The tree shows an ancestral sequence of 20 nucleotides. Along the branches the changes are shown. Deduce the sequences of the five living species.



14. The tree shown for Questions 6–7 has one branch with two trait changes shown, beak long and eating grubs. Why should one be careful about assuming, on the basis of this diagram, that the diet trait changed before the beak trait? What additional information might support the conclusion that indeed beaks lengthened after the transition to grub eating?

15. Trait evolution is often represented with a line drawn across a branch of a phylogenetic tree. Describe the evolutionary phenomena that are represented by this line.

16. Dogs have a tiny appendage above their paw, the so-called dewclaw. If we had a time machine, how could we determine definitively whether this body part is homologous to the human thumb? Assume that if we look at a parent and its offspring we can unambiguously identify the structures in each that are homologous.

CHAPTER FIVE

Relatedness and Taxonomy

Following the acceptance of evolutionary theory in the nineteenth century, ***taxonomy***, the science of establishing and using biological classifications, acquired a clear mission to represent evolutionary relationships. In so doing taxonomy became a branch of ***systematics***, the study of the evolution of biological diversity. Nonetheless, it took another hundred years for all of the implications of that merger to become clear. First, taxonomists had to clarify that evolutionary relatedness should be defined in terms of common ancestry. Second, because common ancestry is captured in phylogenetic trees, taxonomists had to adapt to the idea that their classification systems should mirror the tree of life. That is to say, classification should follow the rule that the more recently two organisms last shared a common ancestor, the more closely they should be classified. And third, systematics needed to develop the tools to build reliable trees. In short, it took biologists over a hundred years to recognize that classification, relatedness, and trees are really three sides of the same coin, so to speak.

In this chapter, we clarify the concept of relatedness and show that it applies equally to family trees (pedigrees) and phylogenetic trees. Then we discuss the common confusion between relatedness and similarity and explain why modern taxonomy focuses on phylogenetic relatedness rather than similarity. We explore ***nomenclature***, the rules regulating the names of taxa, and show why the concept of taxonomic rank is not meaningful in a phylogenetic context. We end the chapter with a brief discussion of phylogenetic nomenclature, a new, but controversial, approach that attaches names to clades rather than to ranked taxa.

THE CONCEPT OF RELATEDNESS

The concept of relatedness as applied to the branches of a phylogenetic tree mirrors that used in discussions of human familial relationships. In both cases, the

degree of relatedness of two living organisms is dictated by how many generations earlier they last shared a common ancestor. Thus, the following two statements should be interpreted in parallel: (1) you are more closely related to your siblings than to your first cousins, and (2) you are more closely related to chimpanzees than to mice.

What does it mean to say you are more closely related to your siblings than to your first cousins? When we pose this question to classes, the most common first response is that you share more genes in common with your siblings than with your first cousin. While this is probably the case, is it the true basis of the closer relationship or just a consequence of the closer relationship? Suppose that through a freak of genetic segregation you actually shared more genes in common with your first cousin than with a sister (we will leave you to work through how this improbable event might happen), would you now say that you are more closely related to your first cousin than to your sister? No, of course not. Relatedness is about kinship and ancestry, not about the actual assortment of genetic material.

The real measure of relatedness is not genetic similarity, nor any other kind of similarity, but recency of common ancestry. You are more closely related to a sibling than to a first cousin because you share more recent common ancestors with your sibling (your parents) than you do with your first cousin (your grandparents). Likewise, you share more recent common ancestors with your

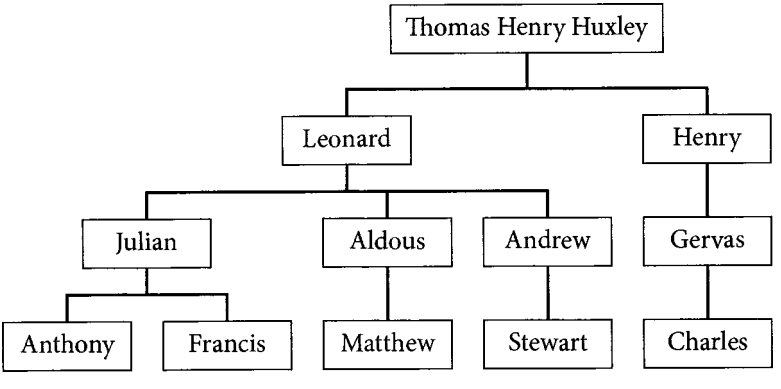


FIGURE 5.1 The pedigree of some male descendants of Thomas Henry Huxley. Four generations of Huxleys are shown with individuals in the same horizontal row (e.g. Leonard and Henry) belonging to the same generation. The oldest generation is at the top and the most recent at the bottom.

first cousins (grandparents), than you do with your second cousins (great grandparents). And so on.

As an example, Figure 5.1 shows some of the male descendants of the famous biologist and Darwin contemporary, Thomas Henry Huxley. By focusing only on the male line, the pedigree is treelike. You could equally follow the female line to obtain a treelike pedigree. Chapter 6 considers more complicated examples where we track both the maternal and paternal lines.

Let us focus on individuals in the lowermost (most recent) generation, analogous to living species. This diagram includes Anthony, his brother (Francis), their two male first cousins (Matthew and Stewart), and one of their (perhaps many) male second cousins (Charles). Anthony is more closely related to Francis than to anybody else in his generation because they share the most recent common ancestor (their father Julian Huxley, who happens also to have been an evolutionary biologist). Likewise, Anthony is more closely related to his first cousins, Matthew and Stewart, than to his second cousin, Charles because Anthony shares a more recent common ancestor with Matthew and Stewart than he does with Charles. The last common ancestor he shares with his first cousins is their grandfather (Leonard), but the last common ancestor he shares with his second cousin is their great-grandfather (Thomas Henry).

You may have noticed that the names of the parents and grandparents are not needed to determine degree of relatedness—all that matters are the lines of descent. Figure 5.2, therefore, provides all the relevant information on the degree of relatedness between members of this generation. The polytomy in the tree results from the fact that Leonard fathered three sons. Given this tree, some

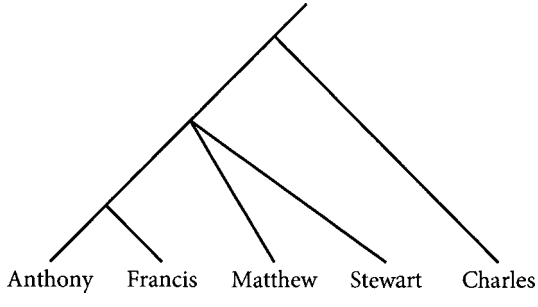


FIGURE 5.2 A tree representation of the relatedness of the five male descendants of Thomas Huxley listed in Figure 5.1. All the individuals shown belong to the same generation. Their relationships may be determined despite the fact that common ancestors are not named.

three-way statements of relationships are true, for example, "Stewart is more closely related to Anthony than to Charles," and others are false, for example, "Francis is more closely related to Matthew than to Anthony." You might find it useful to list and score as true or false, some of the 60 possible three-way statements of relationship involving the five individuals in the most recent generation.

RELATEDNESS AND PHYLOGENETIC TREES

Now we apply exactly the same principle to species relationships. Why is it true to say that a human is more closely related to a chimpanzee than to a mouse? The answer is given in Table 5.1, which shows that humans share a more recent common ancestor with a chimpanzee (ca. 6 Ma) than with a mouse (ca. 75 Ma). Similarly, humans share a more recent common ancestor with a mouse than with a frog (ca. 350 Ma). Thus humans are more closely related to mice than to frogs.

Given that the times that common ancestors lived can be difficult to determine, and also difficult to memorize en masse, it is fortunate that all the information needed to evaluate the relative degree of relatedness is present in a tree diagram. Figure 5.3 summarizes the relationships without needing to include actual dates.

Reading relationships from trees is straightforward provided that you remember that relatedness is defined in terms of recency of common ancestry. As an example, we can use Figure 5.4 to assess whether a salamander is more closely related to a human or a lungfish. The correct answer is that a salaman-

TABLE 5.1 The approximate divergence times between humans and other species.

	Last common ancestry with human
Frog	350 Ma
Mouse	75 Ma
Chimpanzee	6 Ma

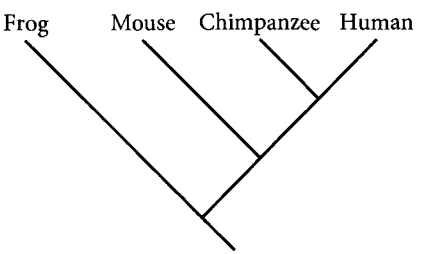


FIGURE 5.3 **Phylogenetic tree showing the relationships among humans, chimps, mice, and frogs.**

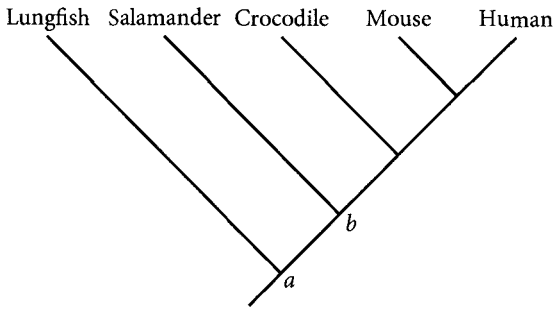


FIGURE 5.4 Phylogenetic tree of five animal species. Node *a* represents the last common ancestor of all the tips (the root node). Node *b* represents the last common ancestor of salamander, crocodile, mouse, and human.

der is more closely related to a human than to a lungfish, because it shares a more recent common ancestor with a human than with a lungfish.

The best strategy for correctly answering the question is to work down the tree to find the point that corresponds to the most recent common ancestor of a salamander and a lungfish. This ancestor is situated at the node labeled *a*. Likewise, the last common ancestor of a salamander and a human is at the node labeled *b*. Because, *b* is a descendant of *a* (to get from the root to *b* you need to pass through *a*), *b* must have lived after *a*. Therefore, a salamander is more closely related to a human than to lungfish.

Interestingly, a significant number of students, and even some professional biologists, reason that a salamander is more closely related to a lungfish than to a human. Research has identified several sources of confusion about trees that combine to explain the high frequency of mistakes in questions of this sort. These issues are worth exploring as a way to make explicit some important aspects of tree thinking.

One common error is to look "along" the tips, and focus on the proximity of labels to one another. This converts the tree into an ordered list: lungfish-salamander-crocodile-mouse-human. Because the salamander is directly next to the lungfish but relatively far away from the human, it may seem to be more closely related to the lungfish. Reading across the tips will often lead to incorrect conclusions about relatedness.

To see why the ordering of tips is not a good guide to relationships, recall that it is arbitrary which descendant lineage is drawn to the right or left. Thus, two trees with different tip orders can tell the same history (see Chapter 3). For

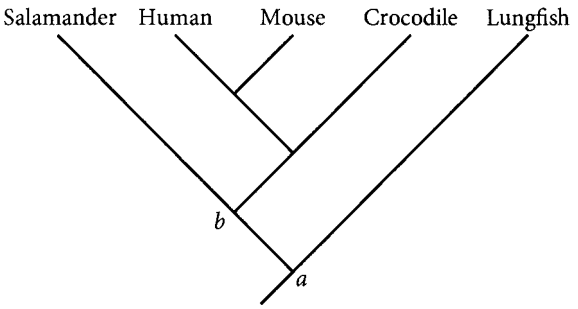


FIGURE 5.5 Phylogenetic tree of five animal species with branches rotated relative to Figure 5.4. Despite this rotation, the relationships among the tips are unchanged. Once again, node *a* represents the last common ancestor of all the tips, and node *b* represents the last common ancestor of salamander, crocodile, mouse, and human.

example, Figure 5.5 shows one of the many ways of rearranging the tips on Figure 5.4 without changing either the tree or the implied relationships. Although the tips have been reordered, the tree topology remains unchanged: *b* (the last common ancestor of salamanders and humans) is still a descendant of *a* (the last common ancestor of salamanders and lungfish).

Conversely, two trees with the same ordering of tips can imply quite different relationships. In Figure 5.6 the bottom tree has the tip order retained from the original tree, but the branching pattern shows some truly bizarre evolutionary relationships. For example, on this tree a mouse is more closely related to a crocodile than to a human. If this messed-up tree *were* correct, a salamander would indeed be more closely related to a lungfish (ancestor *a*) than to a human (ancestor *b*). But it is not!

A second common error arises from "node counting." This is where one notices that to go from the salamander tip to the lungfish tip one need pass only two internal nodes: *a* and *b*. In contrast, to get from the salamander to the human one needs to pass through three nodes: *b* and the two nodes above. By this reasoning you might incorrectly conclude that a salamander is separated by fewer lineage splitting events from a lungfish than from a human, meaning that it is, in some sense, "closer" to a lungfish.

Remember that the form of a tree does not change when tips are pruned off (Chapter 3). So, the number of taxa that are included and happen to have nodes on the path between salamander and human is variable and uninformative. Consider the two trees in Figure 5.7. Both imply the same relationships of

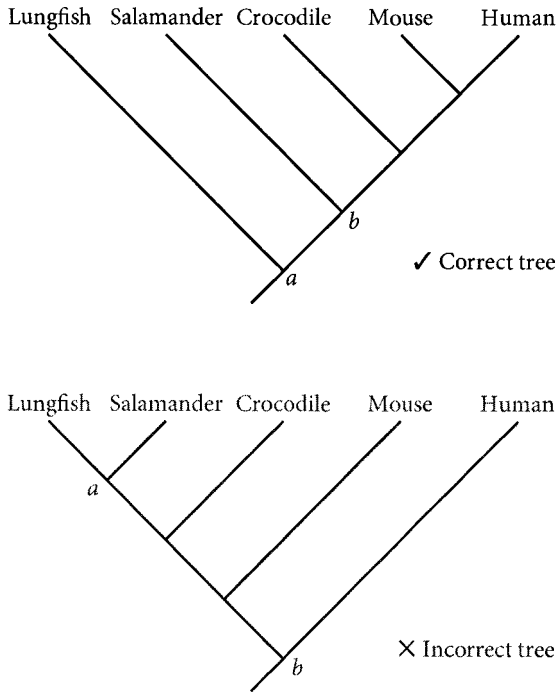


FIGURE 5.6 Tip order is a poor guide to relatedness. Both trees have the same tip order but only the top tree shows the correct relationships.

salamanders to humans and lungfish (notice that node *b* is a descendent of *a* in both cases), but node counting will yield quite different answers.

A third source of confusion arises from the fact that we have (intentionally) chosen species that will likely resonate with popular presentations of evolution as a progressive story with humans as the ultimate target. This ladder of life metaphor has a long history in human culture, only being rejected with the acceptance of the Darwinian evolutionary model (see Chapter 2). Within the ladder framework, you might visualize lungfish giving rise to salamanders, giving rise to reptiles (like crocodiles), giving rise to "lower" mammals (like mice), and then to the "highest" mammal, humans. Therefore, you might think that a salamander is just one rung up from a lungfish, but three rungs down from a human, and is, therefore, closer to the former.

As presented in Chapter 2, ladder thinking thoroughly misrepresents evolution. Salamanders are alive today and none of them are ancestors of humans

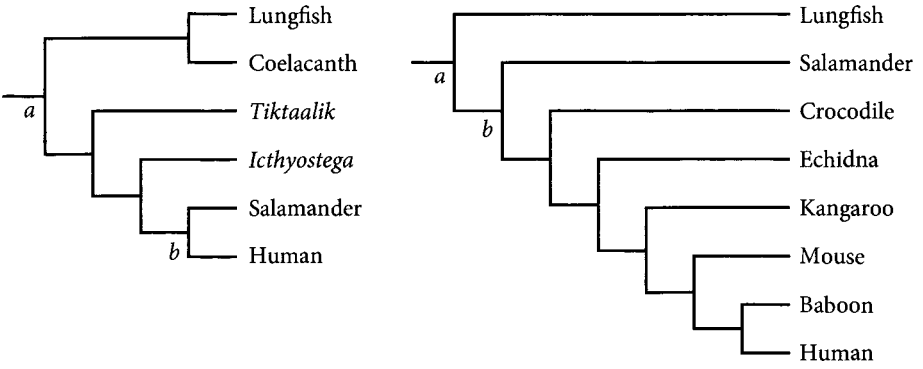


FIGURE 5.7 An illustration that the number of nodes between two tips does not indicate relatedness. In both trees, node *a* represents the last common ancestor of humans and lungfish and node *b* represents the last common ancestor of humans and salamanders. Because *b* is descended from *a*, a salamander is more closely related to humans than to lungfish on both trees. The fact that, on the second tree, salamanders and lungfish are separated by fewer nodes than salamander and humans is irrelevant.

(or any of the other tips). While it is likely that the last common ancestor of a salamander and a human looked (to a human eye, though maybe not to a salamander's) rather more like a salamander than a human, it was neither a living human nor a living salamander. The last common ancestor of you and your first cousin is your grandparent. Even in the improbable case that your cousin looked *exactly* like your grandmother, you would not be descended from your cousin. Relatedness among living species is about lines of ancestry and descent, not about the resemblance of living species to common ancestors.

The final common cause of confusion is a natural tendency to equate relatedness and similarity. You might believe that the similarities of a lungfish and a salamander (slimy skin, lack of hair, aquatic reproduction, etc.) are more numerous than the similarities of salamanders and humans (four limbs, lungs). This, in turn, might lead you to conclude that salamanders and lungfish are more closely related to each other than either is to humans. To make this mistake would be to conflate ancestry with similarity, two entirely different concepts.

If the rate of evolution in visible characteristics were the same on all branches of the tree of life and we only considered living tips, then relatedness and similarity would always match. That is to say, if evolutionary time were always counted in units of "change visible to humans" then the degree of relatedness among a pair of living species would be linearly proportional to their degree of "visible similarity."

However, evolutionary change often accumulates unevenly. Let us assume that changes at the molecular level satisfy a molecular clock, meaning that the rate of substitution (changes per site per unit time) is constant (Chapter 11). Even when a molecular clock holds, some evolving lineages will accumulate few morphological changes because most molecular changes are either silent (not changing the phenotype at all) or affect traits that are not readily apparent, such as biochemistry or physiology. At the same time, other lineages may experience rapid morphological evolution because a higher proportion of molecular mutations affect visible morphological features. Furthermore, molecular clocks often do not hold, which means that the rate of molecular evolution is uneven as well. Taken together, it would be foolhardy to assume that the rate of visible morphological evolution will be identical in two descendants of the same ancestor.

The principle of unequal rates of evolution is illustrated by the example in Figure 5.8. The tree on the left shows (correctly) that crocodiles are more closely related to birds than to lizards. But to the human eye a crocodile seems to have many more features in common with a lizard than with a bird. This is because there was much more evolution in visible traits on the lineage leading to birds than on the other branches of this tree.

The phylogram to the right in Figure 5.8 summarizes the situation. The branch lengths here are not "real" but have been scaled based on the authors' perception as to how much "visible-to-human" evolution has occurred on each branch of the tree. The terminal branch leading to birds can be seen to be much longer than its sister lineage, which leads to crocodile. These two branches are the same length in units of time—both terminals include living species that descended from the same common ancestor. However, presumably driven by

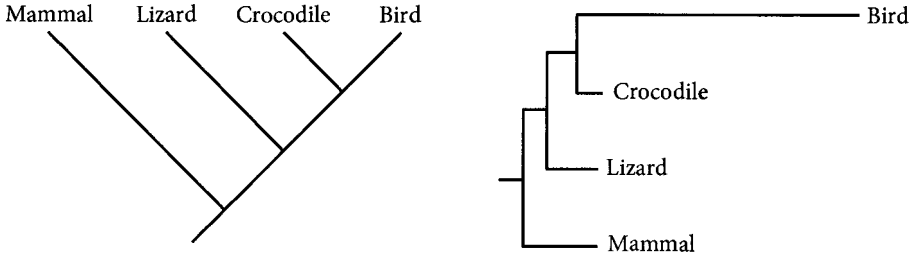


FIGURE 5.8 Cladogram (left) and phylogram (right) for reptiles. In the phylogram, branch lengths have been scaled based on the authors' perception as to the relative amount of "visible-to-human" evolution that has occurred on each branch of the tree. This shows that the bird lineage has accumulated more differences than the crocodile lineage since their split.

the evolution of powered flight, many more physical traits have changed on the bird lineage than on the crocodile lineage. This means that the total number of visible differences that accumulated between a crocodile and a lizard is fewer than the number that separate crocodiles and birds.

The list of the physical traits shared by crocodiles and lizards but absent in birds is longer than the list of traits shared by crocodiles and birds but absent in lizards. In the former category are such features as quadrupedal locomotion, elongated tails, teeth, flat sternum, scaly body, "cold" blood, solid bones, separate collarbones, and many more. In contrast, the latter list would be quite short: for example it would include the behavior of building and defending nests and a skull with an antorbital fenestra (a hole in front of the eye), or AOF. Does the larger number of crocodile-lizard similarities make these two more closely related?

There are two levels at which we need to answer this question. First, do these traits call into doubt the truth of the phylogenetic tree shown? The answer is "no." The features that seem to unite lizards and crocodiles are shared ancestral traits, symplesiomorphies. This is illustrated for the tail trait in Figure 5.9. Non-monophyletic groups of organisms can have shared, ancestral character states (Chapter 4). Thus, these symplesiomorphies do not contradict the claim that crocodiles are more closely related to birds than to lizards.

The traits shared by birds and crocodiles are shared derived traits, synapomorphies. This is illustrated in Figure 5.10 for the AOF trait. However, unlike the shared ancestral states, the shared derived states support this tree by suggesting that crocodiles and birds form a clade (Chapter 4).

Now that we have established that the numerous shared traits of crocodiles and lizards do not challenge the veracity of the tree, we can ask the second ques-

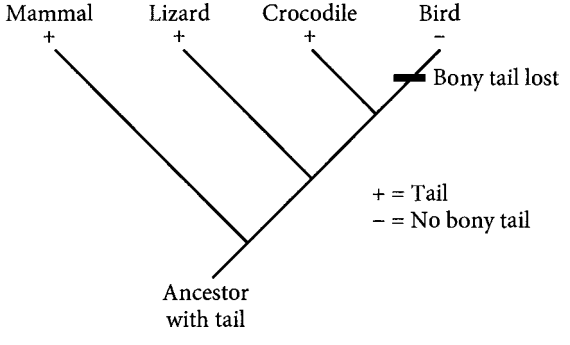


FIGURE 5.9 Tree showing the evolution of bony tail presence/absence in amniotes.

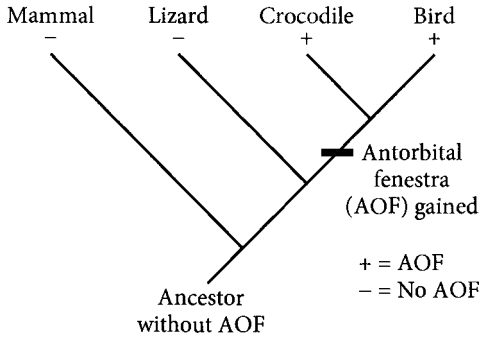


FIGURE 5.10 Tree showing the evolution of the antorbital fenestra (AOF) in amniotes.

tion. Accepting the tree, do the many similarities of lizards and crocodiles make them more closely *related* to each other than to birds? The answer is a resounding "no." The tree depicts common ancestry, the very basis of "degree of relatedness." It would not matter how many derived traits evolved on the bird lineage, the crocodile lineage would still share a more recent common ancestor with birds than with lizards. Relatedness is defined by descent, not similarity.

TAXONOMY AND PHYLOGENY

Taxonomy, the building and application of biological classifications, is an important scientific discipline. By associating groups of organisms with scientific names, taxonomy enhances clear communication about biological diversity. In order to understand why modern taxonomy is so focused on phylogenetic relatedness, it will be useful to review some of the major conceptual changes in taxonomy over the last 250 years.

In the late eighteenth and early nineteenth centuries, naturalists discovered that if one classified organisms on the basis of their traits, a tidily nested taxonomy usually emerged (see Figure 2.6). For example, the vertebrates could be broken up into those with four legs and lungs (tetrapods), most of which lived on land, and those without limbs, which lived in water (fish). Tetrapods, in turn, could be divided into those with an amphibious life style, and those with an amniotic egg that can develop out of water (amniotes). Amniotes, in turn, could be divided into those with feathers and wings (birds), those with fur and milk (mammals), and those with neither feathers nor fur (reptiles).

The system was cleanly nested: there were no feathered animals that lacked an amniotic egg, no furred animals that lacked limbs, etc. This can be summarized in a Venn diagram (Figure 5.11).

Why is the nested structure of taxonomy significant? For contrast, consider the best efforts to arrive at a hierarchical classification of books in a library. The Dewey Decimal system first divides books into 10 broad categories including, Language (400 series), Natural Sciences & Mathematics (500 series), and Geography & History (900 series). Table 5.2 lists some entries that may be found under these three broad categories. As you can see, the same five geographical areas repeat under each of these disciplines.

The classification of these twenty areas according to the Dewey decimal system is shown in the Venn diagram in Figure 5.12. However, unlike the biological case we could propose an equally valid alternative classification of books, as

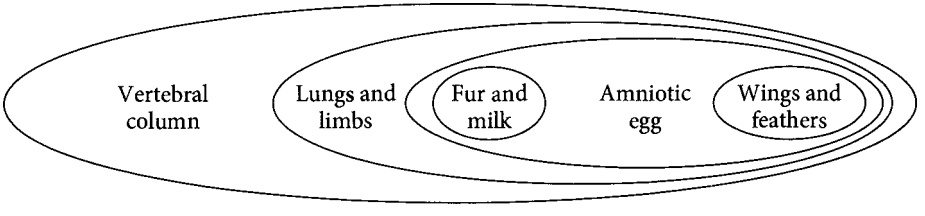


FIGURE 5.11 Venn diagram showing the distribution of certain traits among vertebrate animals.

TABLE 5.2 Dewey Decimal System of library classification

Language		Earth Science		Geography		History	
No.	Topic	No.	Topic	No.	Topic	No.	Topic
420-489	European languages	554	Earth sciences of Europe	914	Geography of Europe	940-949	History of Europe
491-495	Asian languages	555	Earth sciences of Asia	915	Geography of Asia	950-959	History of Asia
496	African languages	556	Earth sciences of Africa	916	Geography of Africa	960-969	History of Africa
497	North American languages	557	Earth sciences of North America	917	Geography of North America	970-979	History of North America
498	South American languages	558	Earth sciences of South America	918	Geography of South America	980-989	History of South America

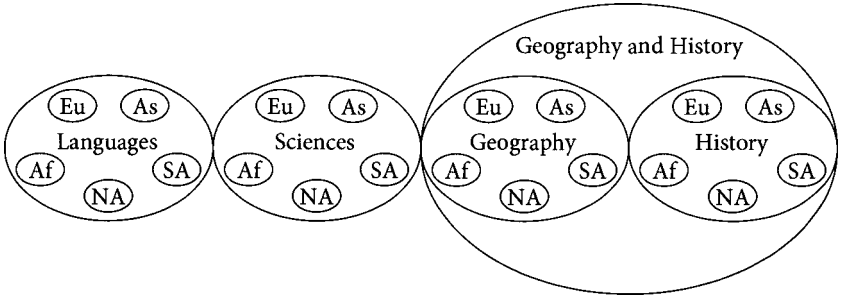


FIGURE 5.12 Venn diagram representation for the part of the Dewey Decimal system in Table 5.2. Af = Africa; As = Asia; Eu = Europe; NA = North America; SA = South America.

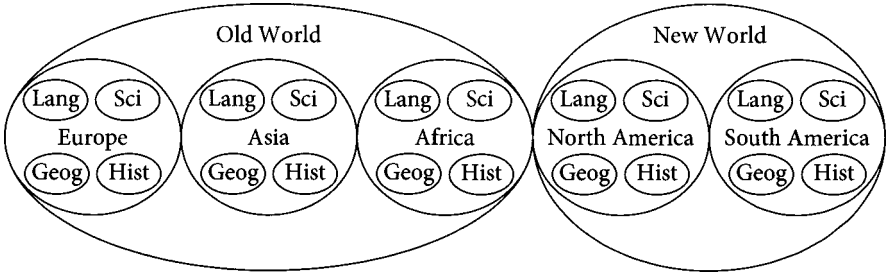


FIGURE 5.13 An equally valid alternative to the book classification used by the Dewey Decimal system. Geog = Geography; Hist = History; Lang = Languages; Sci = Science; NA – North America; SA = South America.

shown in Figure 5.13. Indeed, there are as many equally correct ways of classifying books as there are adjectives to describe books.

The contrast between a classification of convenience, such as that used for library books, and biological taxonomies, was apparent to naturalists before Darwin. It really seemed that there was only one meaningful way to group organisms: classification of books (and the like) is "artificial," whereas that of organisms is "natural." To the pre-Darwinian taxonomist, the only explanation of the naturalness of biological classification was that there was some preexisting structure in God's mind that was represented in His creation.

One of Darwin's great insights was that a nested and hierarchical taxonomy can be explained simply by assuming that the traits used to classify organisms

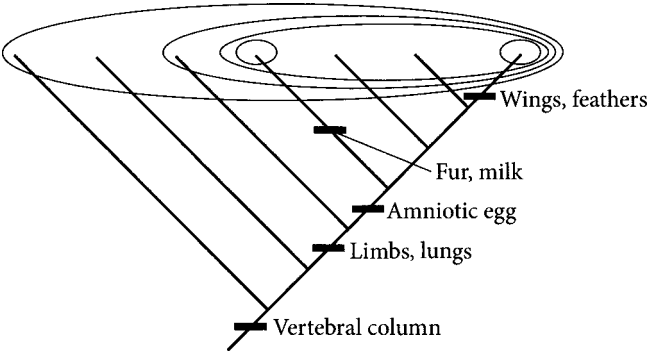


FIGURE 5.14 The phylogenetic history that explains the "naturalness" of the traditional classification of vertebrate animals (as seen in Figure 5.11).

evolved along the branches of an evolutionary tree. For example, consider the following traits: vertebral column, lungs, legs, amniotic egg, feathers, wings, fur, and milk. If we map these traits on the tree of vertebrates (Figure 5.14), as we now understand it, we can see that it perfectly matches the Venn diagram given in Figure 5.11.

Given this phylogenetic tree, we can understand *why* a taxonomy based on these traits yielded a "natural" nested hierarchy. For example, it explains why the amniotes are divided into three non-intersecting groups: mammals, with fur and milk; birds, with wings and feathers, and; reptiles, with neither fur nor feathers. Likewise, it is explicable why all animals with feathers would also have a vertebral column, four limbs, lungs, and an amniotic egg, but wouldn't have fur or milk. If classifications grouped organisms on the basis of their evolutionary kinship, what Darwin called "propinquity of descent" (Darwin, 1859, p. 413), then we *expect* traits to show a nested structure and to yield naturally nested, hierarchical classifications. Thanks to Darwin, the tidiness of biological classification could be explained as being a result of trait evolution along the branches of a phylogenetic tree.

MONOPHYLY AND PHYLOGENETIC SYSTEMATICS

As discussed in Chapter 2, there was a surprising delay between the acceptance of evolution (mid-nineteenth century) and the recognition that taxonomy

needed to reflect evolution (mid-twentieth century). Following Darwin, taxonomists did become more careful to avoid lumping distantly related organisms that shared convergently evolved traits, such as wings in birds and butterflies. However, until the development of phylogenetic systematics, classifications were still built around "important" traits, even at the expense of accurately reflecting evolutionary relatedness. For example, despite the recognition that the vertebrate phylogeny resembles Figure 5.15, vertebrates remained divided into five classes that do not accurately reflect evolutionary relationships.

This classification seemed sound: each of the classes is readily distinguished from the others by one or a combination of traits. Only one trait, limbs, shows homoplasy, and this is not problematic because all groups that have lost limbs have close relatives that have retained limbs. Table 5.3 summarizes the features of the five classes.

This classification of vertebrates into five classes is compatible with these traits, but we now understand that it misrepresents evolutionary relatedness. Specifically, the taxa Pisces, Amphibia, and Reptilia are non-monophyletic (Figure 5.15). To remind you, a monophyletic group, or clade, is composed of all the descendants of a single ancestor (see Chapter 3).

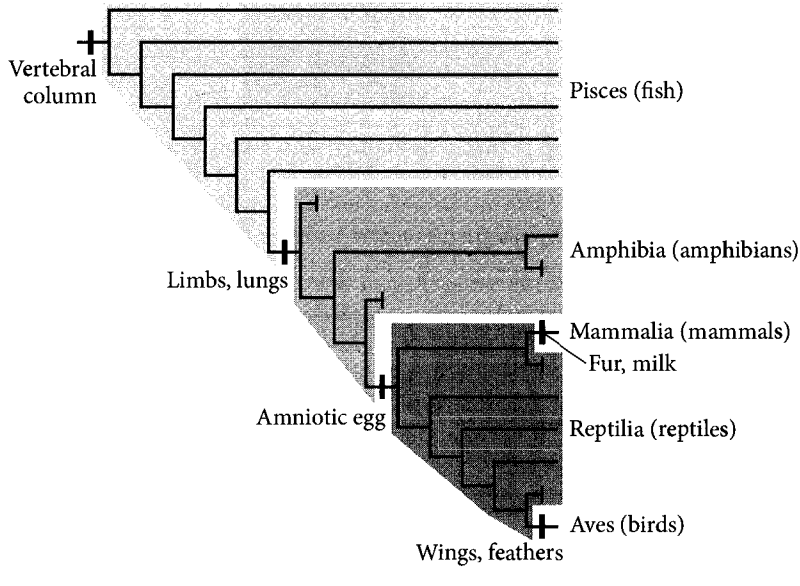


FIGURE 5.15 The traditional classification of vertebrates is based partly on plesiomorphic characters, which explains why three of the five traditional classes are non-monophyletic.

TABLE 5.3 Traditionally the vertebrate classes were defined based on combinations of morphological traits.

	Vertebral column	Limbs & lungs	Amniotic egg	Wings & feathers	Hair & milk
Fish	√				
Amphibians	√	√			
Reptiles	√	√	√		
Birds	√	√	√	√	
Mammals	√	√	√		√

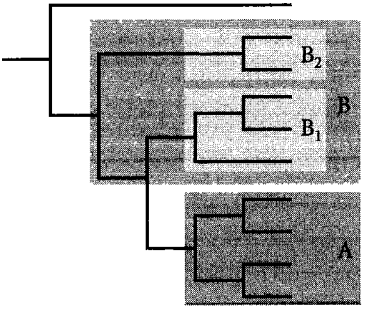


FIGURE 5.16 **The relationship between monophyly and exclusivity.** Non-monophyletic groups such as B, are not exclusive, because some members (e.g., B_1) are more closely related to members of other groups (e.g., A) than to other members of the group (e.g., B_2). Provided there is a strict treelike structure, monophyletic groups such as A, always show exclusivity.

Monophyletic groups have the property that members of the group are more closely related to each other than to any organisms outside the group, a property called ***exclusivity***. As shown in Figure 5.16, members of clade A are all more closely related to the other members of clade A than they are to any organism outside clade A.

Non-monophyletic groups do not show exclusivity: some component organisms are more closely related to organisms outside the group than they are to at least some other group members. For example, in Figure 5.16 the non-monophyletic group labeled B includes some organisms (in clade B_1) that are

more closely related to organisms in clade A than other organisms in group B (specifically, clade B_2). If you consider the traditional vertebrate classes (Figure 5.15) you will see that Pisces are non-exclusive. For example, a lungfish is more closely related to all the land vertebrates (the tetrapods) than it is to other members of Pisces.

From this example, we can induce the general principle that monophyletic taxa (clades), and only monophyletic taxa, have the property of exclusivity. As a result, only monophyletic taxa accurately reflect evolutionary relationships, which is why modern classifications assume that all taxa are monophyletic.

When traditional names are found to apply to non-monophyletic groups, taxonomists generally modify the taxon's content so as to achieve monophyly. This can be illustrated by considering the fate of the non-monophyletic vertebrate classes in modern, phylogenetic classifications (Figure 5.17). Reptilia was contracted when fossil groups formerly placed in Reptilia, but actually more closely related to mammals, were removed. At the same time, Reptilia was

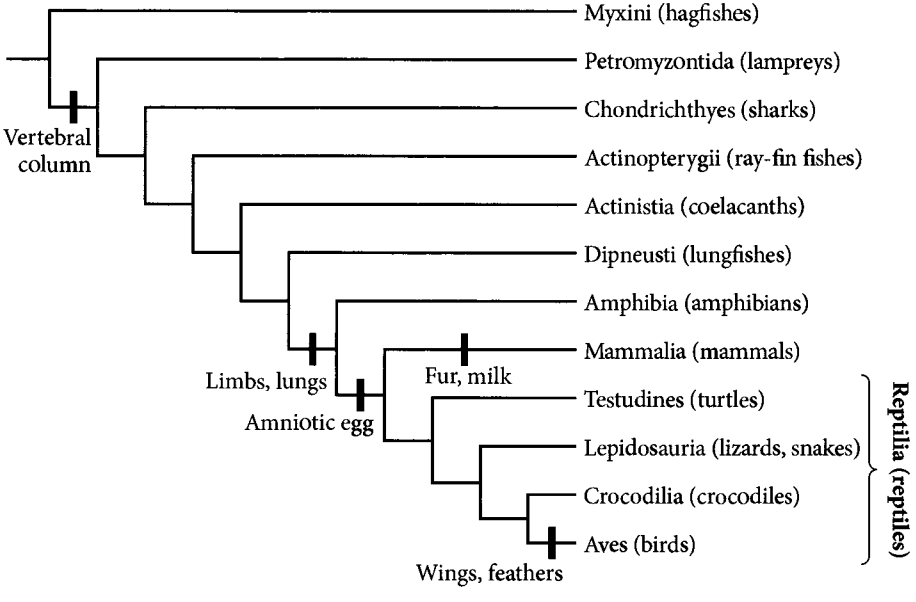


FIGURE 5.17 **A modern phylogenetic taxonomy of vertebrates is consistent with the inferred phylogeny.** This systematization is fully phylogenetic in that only monophyletic groups are recognized and named. Modified from Hickman et al. (2011).

expanded to include Aves (birds) and other organisms that are more closely related to lizards and snakes than to mammals. Hence, the taxon Reptilia now includes the living groups Testudines (turtles and tortoises), Lepidosauria (tuatara, lizards, and snakes), Crocodilia (crocodiles), and Aves (birds). Reptilia also includes many fossil groups, for example pterosaurs, ichthyosaurs, and terrestrial dinosaurs.

The taxon Amphibia, was simply contracted to exclude any extinct taxa that are ancestral to amniotes. There has been no change in the living taxa assigned to Amphibia (frogs, salamanders, and caecilians) given that they happen to form a clade relative to all other living tetrapods.

Finally, the name "Pisces" (fish) was discarded as a scientific name. Instead, modern taxonomies (Figure 5.17) recognize a number of smaller groups with living members, each of which is monophyletic (the relationships of coelacanths and lungfish are still uncertain). The term "fish" has now lost all taxonomic meaning and instead refers to assorted lineages of aquatic vertebrates that have no land-dwelling ancestors.

VARIETIES OF NON-MONOPHYLY

It is common to distinguish two varieties of non-monophyly: paraphyly and polyphyly. Because you will encounter these terms in scientific publications, a brief clarification is required.

Paraphyletic groups include the most recent common ancestor of all group members and some, but not all, descendants of that ancestor. Paraphyletic groups are ones that were formerly recognized as taxa because of shared ancestral character states (Figure 5.18). Fish, amphibians, and reptiles, in their old usage, are paraphyletic groups. In each case they are united by the plesiomorphic character states of the clade that includes them, with one or more parts of that clade excluded from the group (because of the possession of an apomorphic character state). For example, the traditional class Pisces (as shown in Figure 5.15) comprises vertebrates that have not evolved synapomorphies of the tetrapods (limbs and other traits associated with the invasion of land).

Polyphyletic groups, in contrast, are based on convergently evolved, nonhomologous characters (Figure 5.18). For example, if we grouped birds and mammals in a taxon called Homeothermia because they share possession of a homeothermic metabolism (so-called "warm blood"), we would produce a

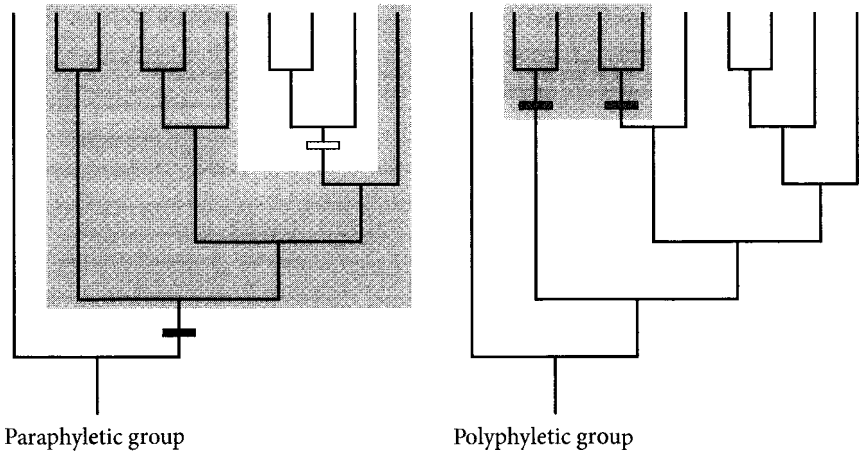


FIGURE 5.18 **The distinction between paraphyly and polyphyly.** A paraphyletic group is characterized by shared ancestral character states. For example, in this figure the paraphyletic group is diagnosed by the presence of a plesiomorphic character (black bar) but lacking a derived character (white bar). Thus, a paraphyletic group includes an ancestral lineage and some, but not all, of its descendants. A polyphyletic group is one united by similar, but non-homologous traits (grey bars) that result from convergent evolution. A polyphyletic group does not include the common ancestor of all of its members.

polyphyletic group. Because homeothermy evolved separately in the bird and mammal lineages, the defining trait (homeothermy) is not homologous among all members of Homeothermia.

The distinction between paraphyly and polyphyly, while commonly emphasized, is not very meaningful from a modern, tree-thinking perspective. Given tree thinking, taxa are defined by common ancestry not by shared traits. While traits are important sources of evidence for the composition of groups, and can often be used to diagnose group membership, they do not *define* group content. Tetrapoda includes all organisms in a particular clade of organisms, even if some of those organisms lack the diagnostic trait of having four limbs.

Given this perspective, paraphyletic and polyphyletic groups have the same basic problem: they are composed of organisms that are not united by uniquely shared common ancestry. The difference between paraphyly and polyphyly boils down to the nature of the mistake that was made when taxa were historically recognized. Paraphyletic groups result from the error of delimiting taxa based on shared ancestral traits, whereas polyphyletic groups result from the

error of delimiting taxa based on convergently evolved traits. Given this, we prefer the general term non-monophyly except in cases where one is explaining the source of error in a traditional taxonomic scheme.

CONVERTING MONOPHYLETIC TAXONOMIES INTO TREES

Because a tree is composed of a set of nested clades, a monophyletic taxonomy will be perfectly consistent with the tree upon which it is based. This means that we can always represent a hierarchical taxonomy in tree form. Being able to interconvert textual or graphic representations of a taxonomy into a tree and vice versa is an important tree-thinking skill. Once developed it becomes possible to take information in a taxonomy and connect it seamlessly to other kinds of evolutionary data.

To convert a tree into an exhaustive taxonomy, you need only list the clades that need to be named and then arrange them in either a Venn diagram or indented-list format (the two most common ways to represent a classification). The Venn format is rather unwieldy in practice, but you can easily see how it mirrors the tree (Figure 5.19), or the Newick (parenthetical) tree format. The indented list format warrants some explanation.

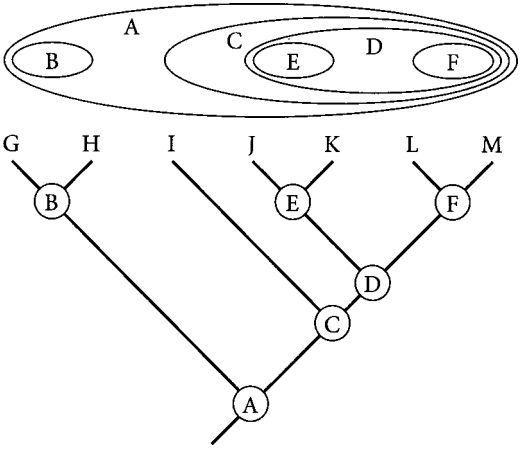


FIGURE 5.19 A tree can be converted into a Venn diagram form. The clades names (A–F) are shown at the node corresponding to their last common ancestor and appear in the corresponding ovals in the Venn diagram. This tree is also represented as an indented classification in Table 5.4.

TABLE 5.4 The indented classification corresponding to Figure 5.20

Level 1	Level 2	Level 3	Level 4	Level 5
Clade A				
	Clade B			
		Species G		
		Species H		
	Clade C			
		Species I		
		Clade D		
			Clade E	
				Species J
				Species K
			Clade F	
				Species L
				Species M

To make an indented list from a tree, we first need to assign a name to each clade. In Figure 5.19, each clade name is shown at the internal node that marks the last common ancestor of the clade. Once we have the clade names we need to build a list in which the names of less-inclusive clades are indented relative to the clades that include them. For example, converting the tree in Figure 5.19 into the indented list format requires the following steps.

1. List the name of the most-inclusive clade: clade A
2. Indent by one tab, and then list the subclades of A: clades B and C
3. Below clade B, indent, and then insert the names of its subclades: species G and H.
4. Below clade C, indent, and then insert its subclades: species I and clade D.
5. Below clade D, indent and then insert its subclades: clade E and clade F.
6. Below clade E, indent and then insert its subclades: species J and K.
7. Below clade F, indent and then insert its subclades: species L and M.

The result of these steps is summarized in Table 5.4. Note that this tree implies this (and only this) indented taxonomy, and this taxonomy implies this (and only this) tree.

The equivalence of tree diagrams, Newick tree descriptions, hierarchical Venn diagrams, and indented classifications, allows scientists to communicate phylogenetic conclusions in a number of different ways. It is worth becoming comfortable with all four methods and how to interconvert them.

TAXONOMIC RANKS

You are probably already aware that biological classifications contain taxa of different ***ranks***: phylum, family, genus, etc. In this section we address the phylogenetic meaning of rank and the role that ranks play in ***biological nomenclature***: the rules that govern the correct names of taxa.

The first development of formal taxonomic ranks goes back to the Swedish botanist Carl Linnaeus (1707–1778). He focused his attention on what he perceived to be the two basic ranks: genera (singular = genus) and species. The genus rank is above the species rank: a single genus can contain multiple species, but a species is assigned only to one genus. In line with his philosophy that genera were kinds created by God while species were variants that arose later through hybridization, Linnaeus proposed a ***binomial*** system of naming. The genus name is a noun and is capitalized. The species name is composed of both the genus name and the species name (or ***epithet***), which is an adjective, written in lowercase. Both names are conventionally italicized or underlined to reflect the fact that they are in Latin. The correct name for the human species is *Homo sapiens*, which may be abbreviated *H. sapiens* (once *Homo* is indicated), but it is never presented as *sapiens* alone.

During the nineteenth century, a richly subdivided taxonomic hierarchy emerged, with an agreed upon set of ranks from the least inclusive (lowest) ranks, species (sometimes subdivided into subspecies and varieties) and genus (sometimes subdivided into sections or series), to the most inclusive (highest) rank, kingdom. Between these extremes were a set of intermediate ranks including, from lowest to highest, tribes, families, orders, classes, and phyla (singular = phylum). When these did not allow enough scope for organizing taxa, intermediate ranks could be indicated (e.g., superfamily, subgenus, infraorder).

Within a nested taxonomy of monophyletic groups, the only information conveyed by taxonomic rank is the nesting of clades: higher ranked taxa con-

tain lower ranked taxa, not the reverse. It is also generally assumed that sister clades are at the same rank. These principles constrain the ranks of clades, but it is always possible to assign all clades to ranks in such a way that this rule is followed. Figure 5.20 illustrates this with an example. You will see that in order for this pattern to be sustained some taxa are ***monotypic***, containing only one taxon of the next lower rank. For example, genus 1 is the sole genus in tribe 1, which is the sole tribe in family 1.

Figure 5.20 shows that ranks can be applied to monophyletic taxa, but that does not answer the question of whether ranks serve any useful purpose. The main role of ranks in traditional nomenclature is to provide a basis for determining the correct name of a taxon. Therefore, to understand ranks, and why they are currently controversial, we need to briefly examine the rules governing the naming of taxa: nomenclature.

The most basic role of taxonomy is to provide a stable tool for scientific communication, by attaching names to taxa. When someone talks about "Lepidoptera," for example, it is important that this term have an unambiguous meaning. Otherwise, we might find ourselves talking at cross-purposes.

During the nineteenth century, as the number of practicing taxonomists increased, and as explorers returned to Europe with collections from around the world, large numbers of new taxa were named and many traditionally recognized taxa were redefined. This produced great instability and confusion in taxonomic circles. As a result, botanists (then including people studying fungi,

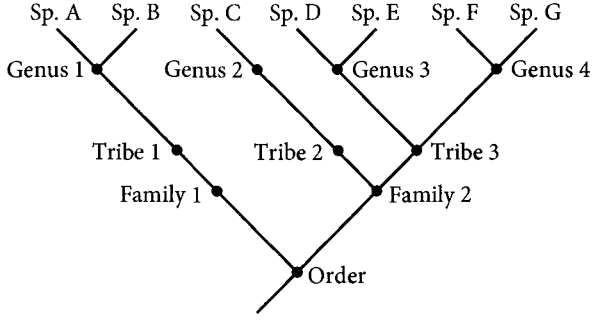


FIGURE 5.20 A ranked classification given a tree. This figure illustrates that all clades can be assigned to ranks in such a way that lower ranked clades are included in higher ranked clades and sister-groups are at the same rank. This requires the creation of monotypic taxa. Sp. = species.

algae, and protists) and zoologists developed formal rules to regulate taxonomic naming (bacteriologists and virologists did so later). The resulting ***codes of nomenclature*** were developed long before phylogenetic principles, or even evolutionary ones, were widely accepted. One manifestation of this disconnect is that these codes heavily emphasize ranks, even though ranks lack clear evolutionary meaning.

The centrality of ranks is best shown by three features of the traditional codes. First, the codes state that all organisms must be assigned to a few mandatory ranks: e.g., species, genus, and family. Second, standardized endings are proposed for many ranks (Table 5.5), meaning that the name of a group changes if its rank changes. Third, and most importantly, the correct name of a taxon is determined by the principle of ***priority at rank***. This states that the correct name for a taxon (defined by reference to particular specimens called ***types***) is the earliest name published at whichever rank it is to be recognized. Priority at rank means that you cannot determine the correct name of a taxon unless you know its rank.

With the development of phylogenetic systematics, taxonomists acquired an objective basis for deciding if a group is a taxon—only monophyletic taxa being considered valid. However to attach names to these clades under the botanical or zoological codes of nomenclature requires that they also be assigned a rank. Is there an objective way to decide on the rank of a clade?

Several ideas have been proposed to try to tie down the concept of rank. One idea is to use the amount of morphological or ecological diversity within a taxon or the degree of phenotypic difference between a clade and its nearest

TABLE 5.5 Some of the main ranks and their prescribed endings in the botanical and zoological codes of nomenclature

Rank	Botanical ending	Zoological ending
Tribe	-eae	-ini
Subfamily	-oideae	-inae
Family	-aceae	-idae
Order	-ales	Not specified
Class	-opsida	Not specified
Phylum	-phyta	Not specified

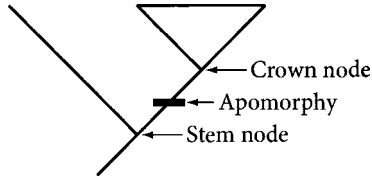


FIGURE 5.21 Events in the history of a clade that can be considered to correspond to the clade's age.

relatives to assign rank. However, there is no objective way to measure phenotypic similarity or difference.

A second option would be to base rank on the number of included species: higher ranks containing more than a threshold number of species. However, the count of species number depends in part on how one defines the species rank, which is unclear. Also, most biologists would balk at the notion that a clade composed of many similar species should automatically be placed at a higher rank than a clade with fewer, but much more diverse, species.

The third and most promising interpretation of rank is based on a group's evolutionary age: older clades being assigned to higher taxonomic ranks than younger clades. However, even ignoring the practical challenges of determining when ancestors lived and when clades acquired their most distinctive traits, this approach has problems. For a start, it is unclear what we mean by "age" of a clade. Does it refer to the age of a clade's ***stem node*** (last common ancestor of the clade and another tip), ***crown node*** (the last common ancestor of all living members of the clade), or apomorphy fixation (the origin of the clade's most distinctive apomorphic trait)? These terms are illustrated in Figure 5.21. Also, defining ranks based on age fails to take account of the varying rates of evolution and diversification in different clades. Thus, even an approach to ranking based on age has failed to be accepted as a universal criterion of taxonomic rank.

PHYLOGENETIC NOMENCLATURE

Given the lack of an objective meaning of rank, it was natural to consider rank-free ways to attach names to clades. In the 1990s, a new system of nomencla-

ture, ***phylogenetic nomenclature***, was developed and gave rise to a new nomenclatural code, the ***PhyloCode***. This system is still controversial and may never become widely adopted, but is nonetheless worth briefly summarizing.

The central idea of phylogenetic nomenclature is that one attaches names to clades rather than ranked taxa. Even if you will never be engaged in nomenclature *per se* it is worth knowing how names can be attached to clades because this provides a useful way to communicate about trees. The two most important clade-naming methods are: node-based and branch-based clade definitions.

A node-based taxon is defined as the least inclusive clade that includes a set of ***internal specifiers***. For example, we might choose to associate the name Mammalia with the least inclusive clade that includes duck-billed platypus, opossum, and human. Each of these specifiers should be associated with a particular specimen in a museum. As shown in Figure 5.22, once we have a tree and know where the specifiers fit, we can immediately determine the clade to which the name Mammalia should apply.

A branch-based taxon is defined as the most inclusive clade that includes the internal specifiers but excludes the ***external specifiers***. For example, Hominoidea might be defined as the most inclusive clade that includes gibbons and humans but not baboons. Hominoidea will, thus, include all organisms that are more closely related to humans (and gibbons) than to baboons (Figure 5.23).

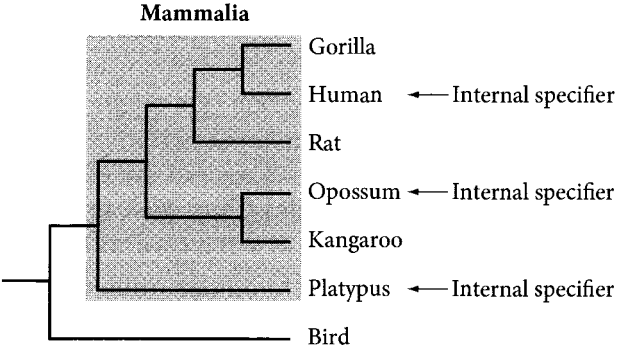


FIGURE 5.22 **A node-based clade definition uses two or more internal specifiers.** For example, suppose that Mammalia were given a node-based definition based on the internal specifiers, human, opossum, and platypus. In that case, given this tree, the grey box marks clade Mammalia.

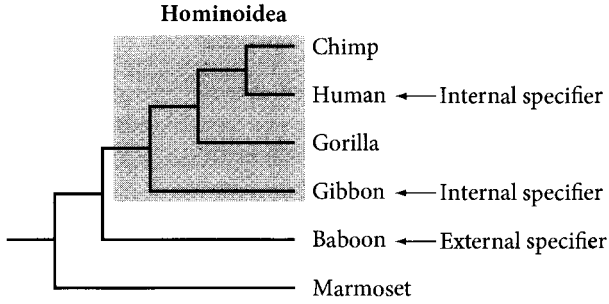


FIGURE 5.23 A branch-based clade definition uses at least one internal and one external specifier. For example, suppose that Hominoidea were given a branch-based definition based on the internal specifiers, human and gibbon, and the external specifier, baboon. In that case, given this tree, the grey box marks clade Hominoidea.

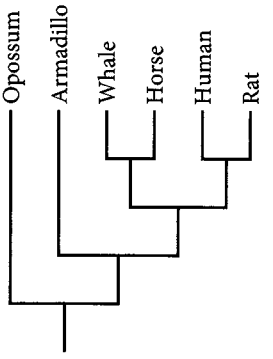
As you can probably see, it ought to be possible to generate a database of phylogenetic names attached to a current best estimate of the tree of life and thereby get rid of a lot of confusion as to the meaning of scientific names. With such an infrastructure, it would be straightforward to update the composition of named clades as information about phylogenetic relationships improved over time. However, as with many such good ideas, the devil is in the details. Thus, while the PhyloCode has many advocates, it remains to be seen whether this, or some other phylogenetic system of nomenclature, will come to replace the traditional rank-based nomenclatural codes.

FURTHER READING

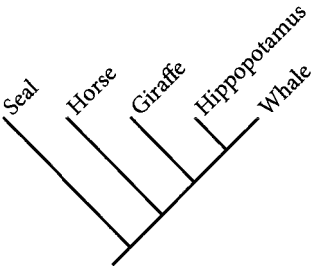
Relatedness: Hennig 1966; Baum and Offner 2008
Taxonomy: Hennig 1966; Hull 1970; Wiley 1981; Podani 2010
Paraphyly/polyphyly: Nelson 1971; Farris 1974; Platnick 1977
Phylogenetic nomenclature: de Queiroz and Gauthier 1990, 1992, 1994; Dubois 2007; Laurin 2008

CHAPTER 5 QUIZ

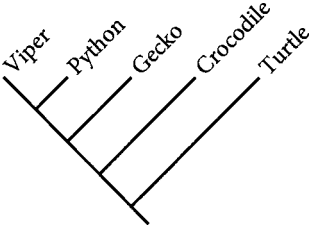
1. Given the tree, which is a correct statement of relationships?
   a. An armadillo is more closely related to a human than to an opossum
   b. An armadillo is more closely related to an opossum than to a human
   c. An armadillo is equally closely related to an opossum and a human
   d. A whale is more closely related to an armadillo than to a human
   e. A whale is equally related to an armadillo and an opossum



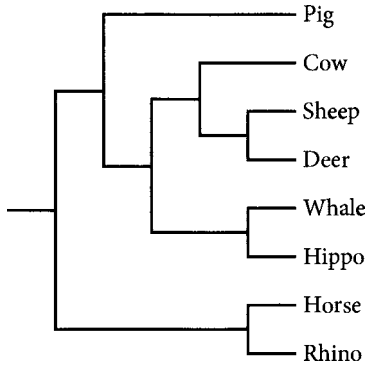
2. By reference to the tree, which of the following is an accurate statement of relationships?
   a. A seal is more closely related to a horse than to a whale
   b. A seal is more closely related to a whale than to a horse
   c. A seal is equally related to a horse and a whale
   d. A seal is related to a whale, but is not related to a horse
   e. A seal is related to neither a whale nor a horse



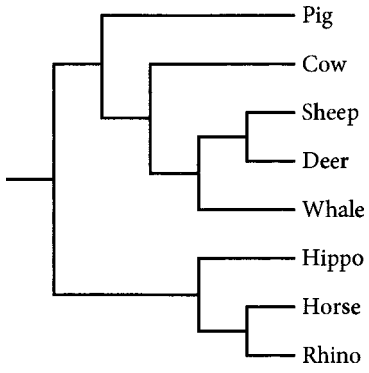
3. Considering the tree, why is a gecko more closely related to a viper than to a crocodile?
   a. Because the gecko and the viper are separated by only two nodes
   b. Because the common ancestor of the gecko and the viper lived before (in the more distant past than) the common ancestor of the gecko and the crocodile
   c. Because the gecko and the viper have a more recent common ancestor than the gecko and the crocodile
   d. Because the gecko and the crocodile have a more recent common ancestor than the gecko and the viper
   e. Because the gecko and the viper are more similar to one another than either is to the crocodile



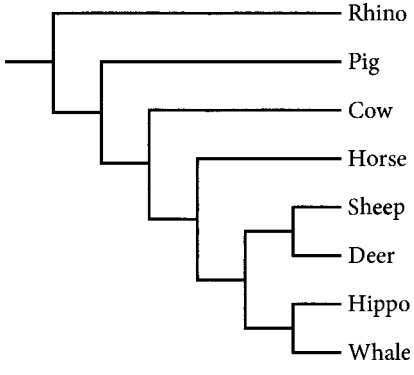
4. Given the following facts, which of the four trees is correct?
   - A whale is more closely related to a deer than to a pig.
   - A cow is more closely related to a pig than to a horse.
   - A hippo is more closely related to a sheep than to a rhino.



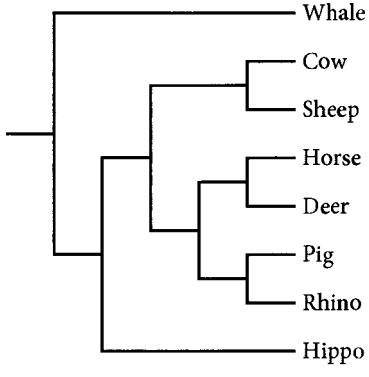
a



b



c



d

5. What is a basic difference between a phylum and a class?
   a. A phylum contains more diversity than a class
   b. A phylum includes more species than a class
   c. A phylum can include one or more classes, but a class cannot contain a phylum
   d. A phylum contains ancestral organisms whereas as class contains advanced organisms

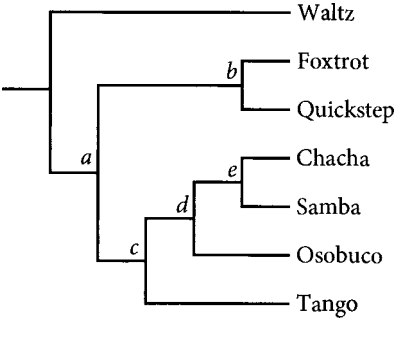
6. The insect family, Curculionidae (the weevils), contains about 40,000 living species, whereas the plant order Amborellales contains just one. Which of the following (if any) can be assumed?
   a. Curculionidae are within the Amborellales
   b. Amborellales are older than Curculionidae
   c. Amborellales is misclassified—the clade should be treated as a genus
   d. Two of the above
   e. None of the above

7. The Tetrapods are a group of vertebrates that includes amphibians, reptiles, and mammals. The last common ancestor of this group had four limbs, hence the name (tetra = four; poda = foot). Is a snake, a reptile that lacks limbs, a member of Tetrapoda, and why/why not?
   a. Yes, because its ancestors had four fins
   b. Yes, because membership is defined by common ancestry
   c. No, because Tetrapoda is *defined* as the group of four limbed animals
   d. No, because snakes are more closely related to non-tetrapods than to tetrapods
   e. It would depend upon whether snakes have limbs as embryos

8. Which of the following is a difference between monophyletic and paraphyletic groups?
   a. Members of monophyletic groups, but not paraphyletic groups, trace to a single ancestor
   b. Members of monophyletic groups, but not paraphyletic groups, can be identified based on homologous traits
   c. Monophyletic groups have shared ancestral characters; paraphyletic groups have shared derived characters
   d. Unlike paraphyletic groups, some members of monophyletic groups are more closely related to species outside the group than to other members of the group.
   e. Unlike paraphyletic groups, all members of a monophyletic group share the same degree of relationship with any species outside the group

9. Which clade corresponds to the node-based definition based on three internal specifiers: Quickstep, Samba, and Foxtrot?

10. Which clade corresponds to the branch-based definition based on the internal specifier Foxtrot and the external specifier Samba?



11. The table shows a classification of the land plants in indented form. The terminal taxa (tips) are not in bold font. The remaining taxa (in bold) correspond to clades. Traits that characterize certain clades are listed on the right. Draw a tree that corresponds to this information with clades labeled and traits marked on the branches on which they evolved.

Taxa	Traits
Embryophyta	Embryo, invasion of land
Liverwort	
Stomatophyta	Stomates
Moss	
Tracheophyta	Vascular system, branching
Clubmoss	Microphylls
Euphyllophyta	-
Moniliformopses	Moniliform spore
Fern	
Horsetail	
Spermatophyta	Seeds
Angiospermae	Flowers, fruit
Grass	
Eudicot	Three-furrowed pollen
Rose	
Pondweed	
Gymnospermae	-
Pine	
Cycad	

12. Convert the tree you created in question 11 into Venn diagram format.

13. It is generally held that a chimpanzee is more similar to a gorilla than it is to a human. However, molecular phylogenetic studies have convincingly shown that a chimpanzee is more closely related to a human than to a gorilla. How can we reconcile these two facts?

14. Traditionally, the "dinosaurs" are considered to have gone extinct at the end of the Cretaceous. Some "dinosaur" groups, for example *Tyrannosaurus,* are more closely related to birds than they are to some other "dinosaurs," for example *Brachiosaurus*. What arguments could be used to suggest that birds are members of Dinosauria and, hence, that dinosaurs are not extinct?

15. Why have some scientists attempted to develop phylogenetic nomenclature, as illustrated by the PhyloCode?

CHAPTER SIX

Gene Trees and Species Trees

The lines of organismal descent that form the tree of life serve as conduits for the passage of genetic material from generation to generation. An ability to visualize the passage of genes through lines of descent is needed to fully understand how traits are transmitted down evolutionary lineages. Furthermore, with gene sequence data emerging as the main tool for reconstructing phylogenetic relationships, it is important to understand the structure of gene histories and how they relate to the histories of populations and species.

This chapter explores the complexities that arise when we consider the ways that genes may be transmitted from generation to generation. This analysis will force us to abandon the simplifying assumption made so far in this book: that there is one true tree relating any given set of tips. As you will see, phenomena such as incomplete lineage sorting, introgression, lateral gene transfer, and gene duplication can result in individual genes tracking trees that differ from the history of the population. These phenomena can cause ***genealogical discordance***, wherein different genes sampled from the same set of tips have different trees. While these phenomena certainly make the interpretation of phylogenetic results more challenging, they also expand the value of phylogenetic data. By looking simultaneously at multiple gene trees, we can gain insights into the evolutionary mechanisms that have acted in the past.

We begin by clarifying the shape of gene trees as they relate to organismal pedigrees and population trees. We move on to examine the effects of gene duplication, and how gene trees can be used to elucidate the history of gene duplication and extinction. Then we consider cases in which population histories are not treelike (due to introgression, hybrid speciation, or lateral gene transfer). This last topic leads naturally into the controversial issue of how the concept of "species" fits into phylogenetic theory.

This chapter, like the preceding five, focuses on developing a theoretical understanding of phylogenetic trees, without worrying about how, in practice, we can reconstruct trees. These six chapters aim to develop your ability to conceptualize trees under the assumption that this will prepare you to more easily grasp the methods that are used to infer trees. That being said, the material covered in this chapter is significantly more challenging than the preceding chapters and is not strictly necessary for understanding methods of phylogenetic inference. Therefore, some readers may find it preferable to skip right to the chapters concerned with phylogenetic inference (Chapters 7–9) before returning to the topics covered here.

GENE TREES IN SEXUAL POPULATIONS

As discussed in Chapter 4, adjacent nucleotides in a DNA sequence will tend to track the same history because they are physically connected to one another. Even though mutational changes at different positions in a DNA strand are independent, they will each be constrained by the same treelike history. In strictly asexual organisms, this logic extends to genes on different chromosomes—because inheritance is uniparental, all nucleotide positions in the genome track the same tree, the tree of organismal descent. Thus, in asexual taxa, genes throughout the genome should have the same true tree topology, and this tree should match that of the organismal pedigree.

In sexual populations things are more complicated. While each nucleotide position in the genome has a strictly treelike history, different nucleotide positions can correspond to different trees. Even if we assume that all the nucleotide positions in a single gene have the same tree, how could different genes in the same set of organisms have different histories?

In sexual organisms each diploid genome has two copies of each gene: there are two alleles at each genetic locus. This usage of the term allele is somewhat different from that introduced in Chapter 4, where alleles were defined as classes of *distinct* gene copies, having at least one difference at the nucleotide level. Here, allele refers to the two copies of a single gene, regardless of how much the copies do or do not differ at the sequence level. Because every allele in every organism has had a distinct evolutionary history (having passed through a different set of ancestral organisms), there is the *possibility* that alleles could have acquired sequence differences, whether or not they actually did so. Thus,

the two meanings of allele (homologous gene copies on different chromosomes; sequence variants segregating in a population) are similar to one another and should not result in confusion.

If we are considering a single locus then all the alleles at that locus are homologous, meaning that they must all trace back to a common ancestral allele (Chapter 4). As a result, they are related through a ***gene tree*** (also called a gene genealogy), which depicts the evolutionary branching events giving rise to the present-day alleles. However, in sexual populations, different loci in the same organisms can have different gene trees. To see why, we need to work through the process of reproduction in a sexual population.

A diploid sexual organism contains two alleles at a genetic locus, one derived from its mother and one derived from its father. Figure 6.1 gives an example of a small sexual population. Each individual is indicated by a pair of circles representing the two alleles. The circle on the left represents the maternal allele, the copy donated by the egg. The circle on the right represents the paternal allele, the copy donated by the sperm. For example, focus on the individual in the top left. Its mother (the source of the left-hand circle) donated her paternal allele, whereas its father donated his maternal allele (as shown by the male and female symbols).

During sexual reproduction in diploids, there are four possible ways that the alleles at a single locus can be passed on. Each offspring could have the maternal or paternal allele from its mother paired with the maternal or paternal allele

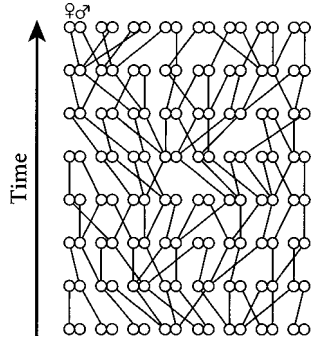


FIGURE 6.1 Lines of descent in a sexual population. Eight generations of a sexual population with eight individuals (paired circles) in each generation. Within each individual, the left circle represents the maternal allele and the right one represents the paternal allele. The lines indicate which allele was passed from parent to offspring.

from its father. With four possible outcomes per mating and, in this example with 56 matings, there are a huge number of possible patterns of genetic descent (9.8 million), of which just one is shown in Figure 6.1.

Despite the fact that the organismal pedigree is netlike, the gene histories are still treelike. To see this, select an allele (a circle) in a generation toward the bottom and move upward to identify all its descendants. We have selected one such allele in Figure 6.2. If you trace the descendants of this ancestral allele, you will see that it forms a tree: that lineages diverge but never converge.

If you picked any two alleles in the current (uppermost) generation, you could trace backward down their lineages until they converged on a common ancestral allele. At the point when the two gene lineages reduce to a single common ancestor, they are said to undergo ***coalescence***. Some alleles from the current generation coalesce only one generation in the past, while other pairs coalesce only at the last common ancestor shared by all alleles in the current generation.

Another way of looking at the same phenomenon is to note that only three of the sixteen alleles present in the lowermost (oldest) generation, marked with arrows in Figure 6.2, have persisted to the present. The others have gone extinct by failure of organisms to reproduce or by a failure of the relevant allele to be

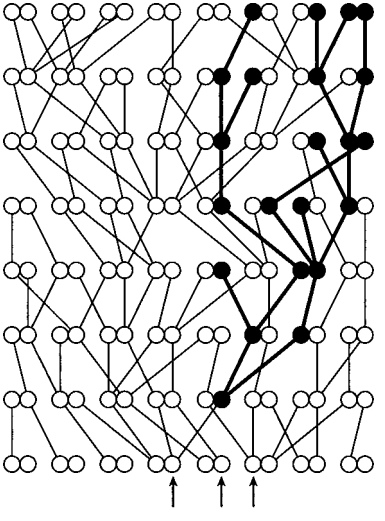


FIGURE 6.2 The treelike history of alleles within a population. All the descendants of one allele in the second generation are shown in black. All alleles in the most recent (uppermost) generation are descended from just three alleles in the earliest generation, marked with arrows.

passed to offspring. The loss of gene lineages over time, ***lineage sorting***, is inevitable. An allele present in one generation can fail to be represented in the next, but in a closed population lineage (with no gene flow) a new allele cannot arise. There is a ratchet: gene lineages can die but cannot be born. This does not necessarily mean that organisms are destined to lose variation. Mutations—while not shown in this diagram—are certain to arise from time to time, meaning that variation lost by lineage sorting is replenished by mutation.

To understand how different alleles in the same population can have different, ***discordant***, gene trees, let us follow two loci on different chromosomes in a sexual population with eight individuals (Figure 6.3). The upper and lower panels of the figure depict the same population. You can tell this by the fact that each organism (pair of circles) is derived from the same mother (source of the left-hand circle) and the same father (source of the right-hand circle). It may be helpful to work through this example carefully to be sure that the commonalities are clear. The only difference between the two independently assorting nucleotide positions is that in about half of the cases a different allele (maternal or paternal) is passed on to the next generation. Focusing on the top left individual, it received its mother's paternal allele and its father's maternal allele in the upper case, but the maternal alleles from both parents in the lower case.

Now, imagine that you sampled six of the eight individuals and looked at one of the two alleles they carry. These alleles are marked with arrows. For each panel we can work down from these six alleles (labeled A–F) to determine the order in which they coalesce. For example, in the upper panel, B and D coalesce three generations before the present, A and C coalesce four generations before the present, and the combined B + D and A + C lineages coalesce five generations before present. This yields a tree with the clade ((A,C)(B,D)), as shown to the right of the figure. Doing the same thing on the lower panel yields a quite different tree.

The basic reason for discordance between these two loci is that chromosomes assort independently during meiosis. This is what allows paternal and maternal traits to be recombined. It is ultimately this recombination that explains why genes within sexual populations can have discordant gene trees.

Because recombination is the cause of discordant trees, the frequency of recombination between two genes determines the probability that their underlying trees will be discordant. At one extreme, genes on different chromosomes are relatively likely to have discordant histories because recombination happens in 50% of the meiotic cell divisions. At the other extreme, two genes that

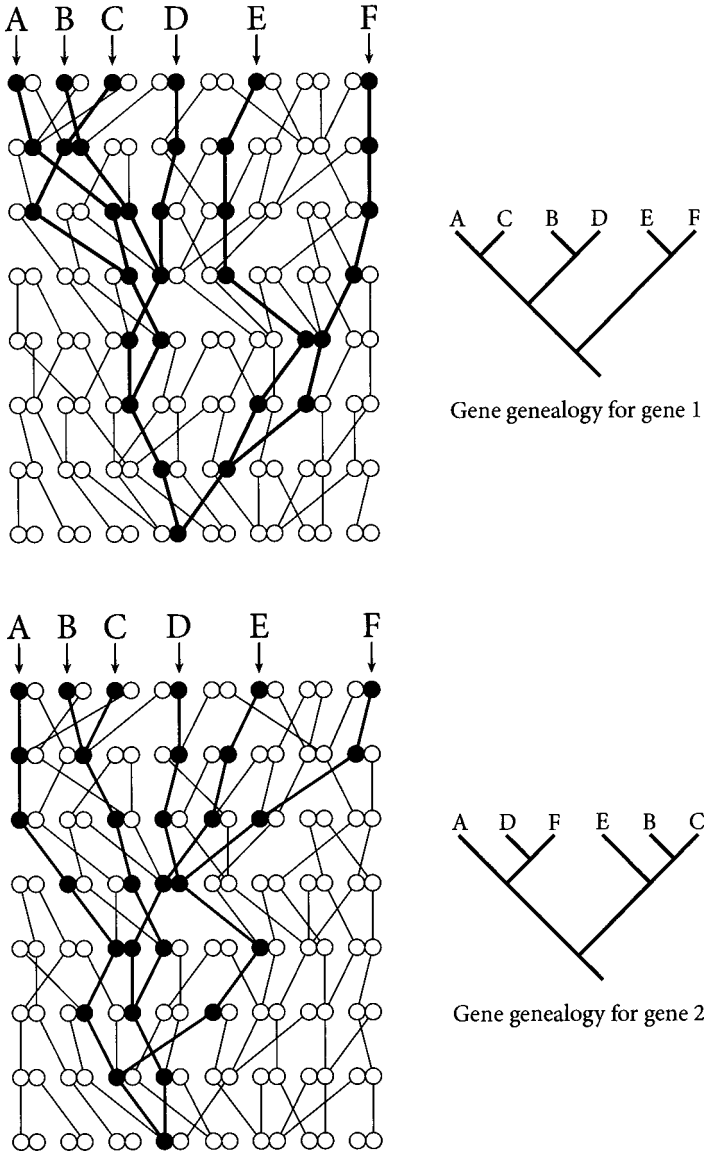


FIGURE 6.3 **Gene-to-gene discordance at two unlinked loci within a single population lineage.** Different coalescent histories for two different loci in the same population. A through F are six individuals; one allele for each was selected to trace the gene histories shown on the right.

are physically adjacent in the genome will be tightly linked: if a gamete gets the maternal allele for the first, it will almost certainly get the maternal allele for the second. Such tightly linked genes will tend to have the same gene tree. Between these two extremes, the degree of discordance is determined by how frequently the two genes tend to cosegregate, which, in turn, is determined by how frequently recombination happens between them. Genes that are close together on a chromosome (i.e., tightly linked) will cosegregate more often than those that are on opposite ends. For more on recombination and linkage, consult a genetics textbook.

Imagine that you had a tree-o-scope, a magical device that you could "touch" to a single homologous nucleotide position in a set of individuals and it would show you the true tree for that position. Suppose that you started at one end of a chromosome and slid the tree-o-scope along the DNA strand, one nucleotide position at a time. What would you see? You would see the same tree for blocks of adjacent nucleotides, but after traveling some distance down the strand the tree could suddenly change. This would indicate that at least one relevant recombination event had occurred. By "relevant" we mean that the recombination event affected the gene tree that links this particular set of individuals. How far along the chromosome you would have to go to see a different history would depend on the recombination rate (and an element of chance).

While we don't have a tree-o-scope, in a few cases where we have complete genome sequence information, we can actually extract data of this sort. Figure 6.4 shows an analysis conducted with three subspecies of house mice (White

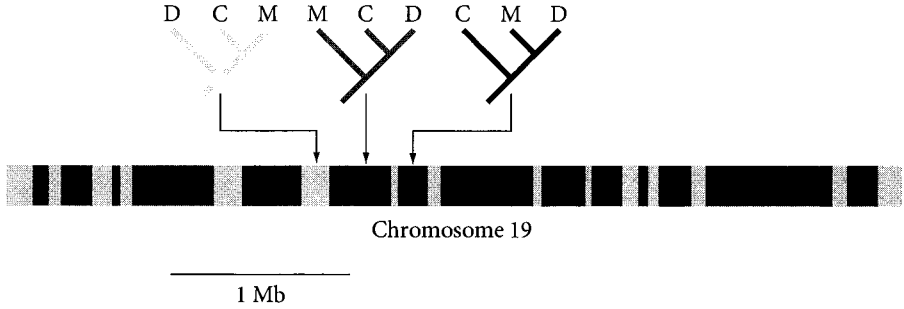


FIGURE 6.4 Variation in gene history along a house mouse chromosome. The shading indicates the tree that applies to each portion of the chromosome. 1Mb = 1million bases. Adapted from White et al. (2009).

et al. 2009). The figure shows the blocks assigned to each of the three possible tree topologies in a stretch of about 5 million base pairs of chromosome 19. All three tree topologies occur, but the tree with "C" and "D" subspecies forming a clade is true of the largest subset of this chromosome.

A collection of adjacent nucleotides that have the same tree is known as a ***recombinational gene***. As you walk along a chromosome, the periods of constancy in the gene trees correspond to recombinational genes and changes in the gene trees correspond to the boundaries between recombinational genes. In the house mouse case, each shaded block along the chromosome is one recombinational gene and the blocks are bounded by relevant recombination events. Along a given chromosome in a given set of organisms, some recombinational genes are long, containing many adjacent nucleotide positions, whereas others are short. There are now analytical methods for systematically identifying the boundaries between recombinational genes. Most of these entail looking for places where a change in gene tree topology helps to explain patterns of variation among the sequences.

In the field of population genetics and phylogenetics, a "gene" is usually equated with a "recombinational gene." A recombinational gene has no necessary relation to a functional gene, a piece of DNA that encodes a protein (or functional RNA molecule). It is possible for a single functional gene to be split among two or more recombinational genes due to recombination events within the functional gene. However, because most phylogenetic studies occur at a scale where recombination within a functional gene is rare, it is standard to use the term "gene" to refer to a recombinational gene, but to assume that entire functional genes behave as single recombinational genes. For the remainder of this book, we will use the term "gene" in this way. A gene tree is the treelike history of a set of alleles or copies of a single (recombinational) gene.

GENE TREES IN BRANCHING POPULATION TREES

A sexual population is a set of organisms living at a certain time that have the potential to interbreed with one another. It is not necessary, indeed usually impossible, for every organism in a population to mate with *every* other member of the population. It is not even necessary that all pairs of organisms in the current generation are able to mate. For example, two males or two females cannot parent a biological offspring. Rather, membership in a population is

defined by the fact that the component organisms have some reasonably high probability of sharing a common descendant in the future.

An unbranching population lineage is composed of a single population at each point in time. These sequential populations are connected into a population lineage. As discussed in Chapter 3, the introduction of a major geographic break in a population lineage can divide a population into two, thereby precipitating lineage splitting. Before the split, there is only one ancestral population. After the split there are two daughter populations, each one composed of organisms that descend from organisms in the common ancestral population. Each of the daughter populations is the beginning of a new population lineage. A series of successive lineage-splitting events can produce a multibranched population tree. Such population trees contain gene trees: a population tree is like a system of electrical conduits through which gene lineages, like individual wires, must pass.

Consider a simple population tree such as that in Figure 6.5. In this case, an initial population lineage-splitting event, split *a*, produced two lineages, one of which yielded a single living population, A, and the other went through a second split, *b*, to yield two living populations, B and C. Imagine sampling one allele from each tip. What gene tree would you expect?

Alleles from populations B and C must coalesce before population split *b*. Similarly, the coalescence of alleles in A and either B or C must predate split *a*. You might be tempted to go one step further and assume that the coalescence of alleles from populations B and C should be between splits *a* and *b*. That is to

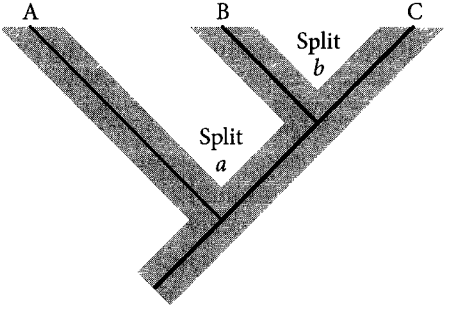


FIGURE 6.5 A gene tree whose topology matches its population tree. Populations (gray lines) form conduits, which contain gene lineages (black lines).

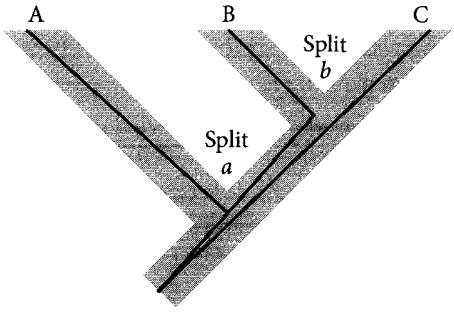


FIGURE 6.6 Gene tree/population tree discordance due to deep coalescence. The gene lineages from B and C fail to coalesce into a single lineage between splits *a* and *b*.

say, you might conclude that the gene tree must match the population tree, as shown in Figure 6.5.

This reasoning is flawed, however. It overlooks the possibility of ***deep coalescence***. A coalescence of alleles in B and C must predate split *b*, but it could also predate split *a*. If it does predate split *a*, then it is possible for gene lineage B to coalesce with gene lineage A more recently than B coalesces with gene lineage C. As a result, the true gene tree can be different from the true population tree. Figure 6.6 shows an example of a gene tree on which an allele from A is sister to an allele from B, despite evolving in a population tree that has a sister-group relationship between populations B and C.

The only way to generate deep coalescence, which can result in a gene tree that contradicts the population tree, is for multiple gene lineages to persist side by side in the population lineage between splits *a* and *b*. In Figure 6.6, every generation between split *a* and *b* contained at least one organism with an allele that is ancestral to B and at least one organism with an allele ancestral to C. To see how this is possible, we need to look at the individual organisms and their genetic compositions.

Figure 6.7 shows a very simple example of a gene history and organismal pedigree that illustrates deep coalescence. Every population contains four individuals. Moving down the tree in this hypothetical history, all of the present-day alleles in each population coalesce to a single ancestral allele before merging with another population lineage. In other words, lineage sorting was complete within the A, B, and C terminal branches.

Now the question becomes, When do the A, B and C gene lineages coalesce? First, notice that between splits *a* and *b*, both the B and C gene lineages were

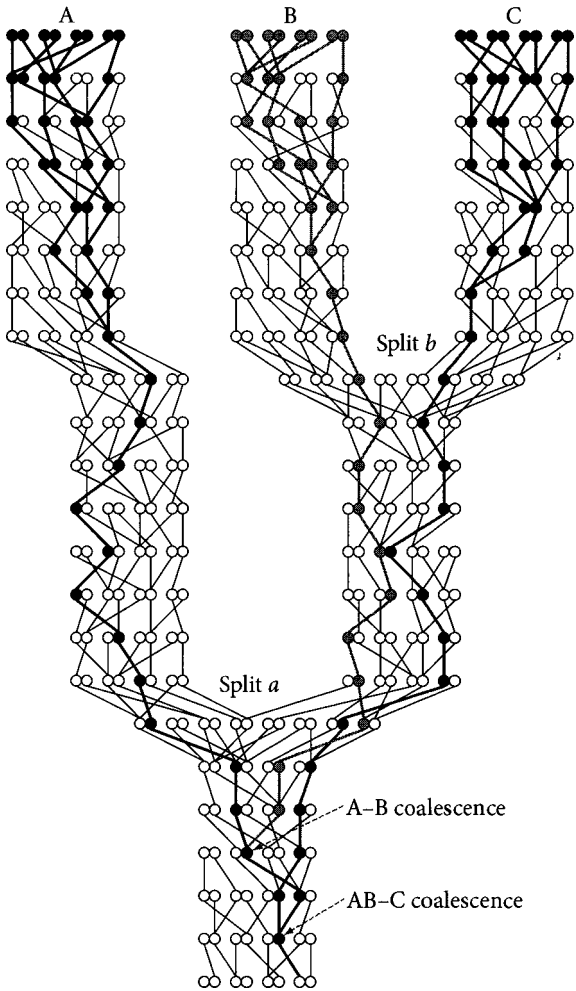


FIGURE 6.7 Detailed view of gene tree/population tree discordance due to incomplete lineage sorting. Deep coalescence resulted in a gene tree ((A,B)C) that is different from the population tree (A(B,C)).

retained in the population: lineage sorting was incomplete. This means that these gene lineages must coalesce prior to split *a*. However, the A gene lineage was also present before split *a*. This opens the possibility that the A lineage coalesced with the B or C lineages closer to the present than they coalesced with each other. Indeed, following the gene histories in this hypothetical case, we see

that A and B coalesce first (working downward), resulting in a gene tree with A more closely related to B than C. Thus, the true gene tree has alleles from organisms in population A and B being more closely related to each other than alleles from organisms in C, even though this topology is at odds with the population history.

This example shows how ***incomplete lineage sorting*** can cause deep coalescence, which in turn can yield conflicts between gene trees and population trees. Because lineage sorting was incomplete between split *b* and split *a* (i.e., the B and C alleles did not coalesce to a single ancestral allele along that branch), the alleles present in A, B, and C could coalesce in an order differing from the order of population splits. Under some simplifying assumptions, we expect one-third of the genes subject to deep coalescence to have gene trees that, by chance, match the population tree, (A(B,C)), whereas the other two-thirds of the gene trees will have one of the two discordant tree topologies: (B(A,C)) or (C(A,B)).

The phenomenon of deep coalescence divides the genome into two pools. Those genes that show complete lineage sorting between split *a* and split *b* will have a gene tree that matches the population tree, (A(B,C)). Those genes that experience incomplete lineage sorting will show deep coalescence, resulting in an equal frequency of (A(B,C)), (B(A,C)), and (C(B,A)) trees. Taken together we expect a plurality or majority of genes to have the (B,C) clade and an equal minority to have either of the discordant gene trees, with an (A,B) or (A,C) clade. The proportion of the genes in these two pools depends on the amount of time separating split *a* and split *b* and the effective population size along this branch (refer to a population genetics textbook).

To illustrate these ideas, consider some recent analyses of genome sequence data from humans, chimpanzees, gorillas, and other primates. In one study, the

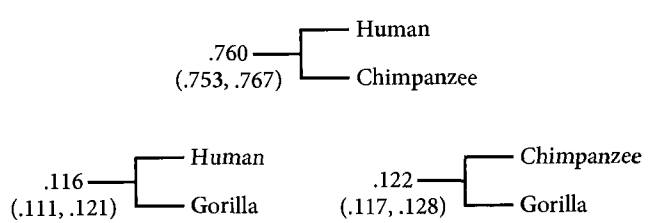


FIGURE 6.8 Genealogical discordance in the human genome. Numbers indicate the proportion of the genome for which the tips shown form a clade (the concordance factor). The numbers in parentheses are 95% credibility intervals around the estimated concordance factors.

genome was divided into recombinational genes, and the proportion of these genes that had each of the three possible resolutions of the human-chimp-gorilla relationship was inferred (Ané 2010). As summarized in Figure 6.8 and discussed more fully in Chapter 9, for 76% of the human genome, humans and chimps form a clade, while the two alternative resolutions each appeared to be true for about 12% of the genes. This pattern is consistent with a population tree with the topology (gorilla(human,chimp)) combined with discordance due to incomplete lineage sorting.

GENE DUPLICATION

As we have just seen, recombination between genes within a sexual population allows genes to have discordant trees. Another possible source of gene-to-gene discordance is gene duplication followed by gene copy extinction.

An entire genome can double such that there are two copies of every gene present, a phenomenon called ***whole genome duplication*** or ***polyploidy***. Alternatively, certain molecular mechanisms can cause a single gene or set of adjacent genes to duplicate independently of the rest of the genome, which is known as ***segmental duplication***. Repeated gene duplication events produce ***gene families***: sets of related genes that have similar, but often somewhat diverged, functions. It is not necessary to go into the molecular mechanisms of duplication, nor the fascinating natural history of gene families. It is important, however, to consider the way that gene duplication shapes gene trees.

Consider the duplication of genes within a single genome. Imagine a gene, A, that exists as a single copy in the haploid genomes of all organisms in an ancestral population (a diploid organism has two haploid genomes and thus two alleles of the gene, but we will ignore this). Suppose that a second exact copy of the original gene is generated somewhere else in the same haploid genome. Let us call the duplicate genes A1 and A2. Gene duplication is a lineage-splitting event in a gene history in which one ancestral gene becomes two descendant genes. Following gene duplication, the two gene copies can accumulate mutations independently and gradually diverge in sequence.

The duplication persists in the long run only if it first arises in an individual that leaves offspring and if eventually it comes to be fixed in a population lineage. Imagine that after being fixed, gene A2 undergoes yet another gene duplication to give rise to genes A2a and A2b. Once this second duplication goes to

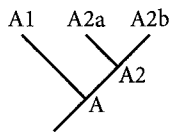


FIGURE 6.9 Two duplication events of a single gene resulting in three copies. Labels at the nodes indicate the name of the clade descended from that node.

fixation, what will be the relationship among the three genes? Since A2a and A2b share a more recent common ancestor with each other than either does with A1, the correct tree is the one shown in Figure 6.9. If we correctly inferred this (rooted) gene tree, we would immediately see that genes A2a and A2b represent a more recent gene duplication, whereas the creation of A1 and A2 was a more ancient gene duplication.

Now we have an opportunity to test your tree-thinking skills: what would the gene tree look like if, after these duplication events happened, the population lineage split to give rise to two living species, X and Y? Species X and Y would each have three gene copies, A1, A2a, and A2b, meaning that there would be six tips. How would they be related to each other?

One way to think through this problem is first to draw the population lineages as though they were hollow tubes. Then you can draw the gene tree inside these tubes, making sure that all gene copies present in an ancestral population make it into the two descendant lineages. Next, you can use tree-thinking skills to "unfold" the gene tree, labeling the genes by the species from which they were sampled. As shown in Figure 6.10, three nodes, each marked with a circle, correspond to the lineage-splitting event (X versus Y), whereas two nodes, marked with squares, correspond to the two gene duplication events.

Before discussing alternative possible histories of gene duplication and lineage splitting, we should clarify some widely used terminology. All genes that descend from a common ancestral gene are homologous genes or ***homologs***. Furthermore, two main kinds of relationships may be recognized for a pair of homologous genes, differing in the role of gene duplication.

If you can track back to the last common ancestor of the two genes without encountering a gene duplication event, and if the last common ancestor gave rise to the two genes via a population lineage-splitting event, the two genes are ***orthologous genes*** or ***orthologs***. For example, XA1 and YA1 are orthologs, because between their last common ancestor (in the common ancestor of X and

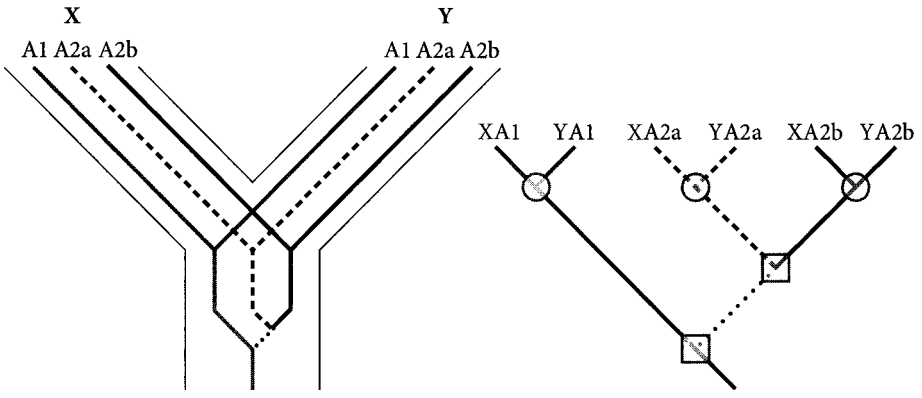


FIGURE 6.10 A lineage-splitting event following two rounds of gene duplication. On the right is the gene tree and on the left is the gene tree folded inside the population tree. Nodes with circles correspond to population lineage splitting events, whereas nodes with squares correspond to gene duplication events.

Y) and the present there have been no gene duplication events. Orthology thus describes the relationship among genes that arise when there has been no relevant gene duplication.

Alternatively, if you need to cross a gene duplication event to get back to the last common ancestor, then the genes are not orthologs. Specifically, if the last common ancestor generated its descendant lineages by gene duplication (and there are no other confounding duplication events), then the genes are ***paralogous genes*** or ***paralogs***. For example, XA1 is paralogous to YA2a because the last common ancestor of these two genes was the root node, which corresponds to the A1-A2 gene duplication event. Paralogy describes the relationship between two genes (maybe in the same genome) that derive from a gene duplication event.

Other variants of gene homology apply when one has to cross gene duplication events before getting back to the last common ancestor. A complex taxonomy of names has been proposed to accommodate these cases. We will not introduce these terms since they are not widely used outside of the field of molecular evolution.

It is worth clarifying that the concepts of orthology and paralogy relate to the cause of the existence of distinct gene lineages: population splitting or gene duplication. They do not refer to the current role that a gene plays in the

development of an organism, that is, its molecular function. Nonetheless, the identification of orthologs is often an important step in functional genetic studies. When looking between species, orthologs have more recent common ancestors than paralogs do. For example, XA1 is more closely related to YA1 than to either YA2a or YA2b. As a result, if the biochemical functions of genes change slowly, a gene in one species is more likely to have a similar genetic function to its ortholog in a second species than to a paralog in the same second species. However, there will be exceptions. For example, supposing that YA1 acquired a novel role in the organism's development, while the other genes retained an ancestral function, XA1 would have a more similar function to YA2a than to YA1. This is another manifestation of the principle that trees depict relationships, not similarity (Chapters 4 and 5).

In practice, the history of gene duplication is usually not known. Nonetheless, and especially when the species tree (i.e., the relationships among the sampled species) is well understood, it is possible to learn about the history of gene duplication and loss by analyzing a gene tree's topology. One simple approach utilizes the principle of parsimony, which was introduced in Chapter 4.

Figure 6.11 gives an example of inferring the history of gene duplications with parsimony when the species tree is known. Imagine we have found that a

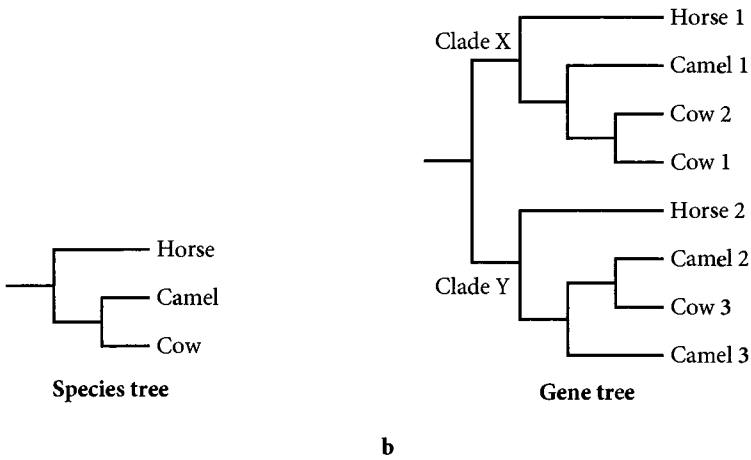


FIGURE 6.11 **Reconciling a gene tree and a species tree.** A hypothetical gene family in ungulates. (a) Relationships of three species: cow, camel, and horse. (b) Gene tree for the hypothetical gene family.

certain gene family contains two gene copies in horses and three each in cows and camels. We are confident that cows and camels are more closely related to each other than to horses (Figure 6.11a) and that the eight genes have the tree shown (Figure 6.11b). The principle of parsimony can be used to reconcile the gene tree with the species tree, minimizing the number of gene duplication and gene deletion events. Computer programs are available for doing these calculations, but with practice one can often generate a plausible hypothesis "by eye." We will walk through this example to illustrate the principles at work.

You will see two major clades on the gene tree, labeled X and Y. Since genes from all three species occur in both of these major clades, we can assume that clades X and Y are the result of an ancient gene duplication that predated the divergence of the three species. This is supported by observing that within clades X and Y the deepest node resembles the species tree in having horse genes on one side and cow plus camel genes on the other. The ancient gene duplication giving rise to clades X and Y is shown in Figure 6.12.

Looking within clade X, you will notice that cow 1 and cow 2 are sister genes. They must have originated at some point via gene duplication. It is most

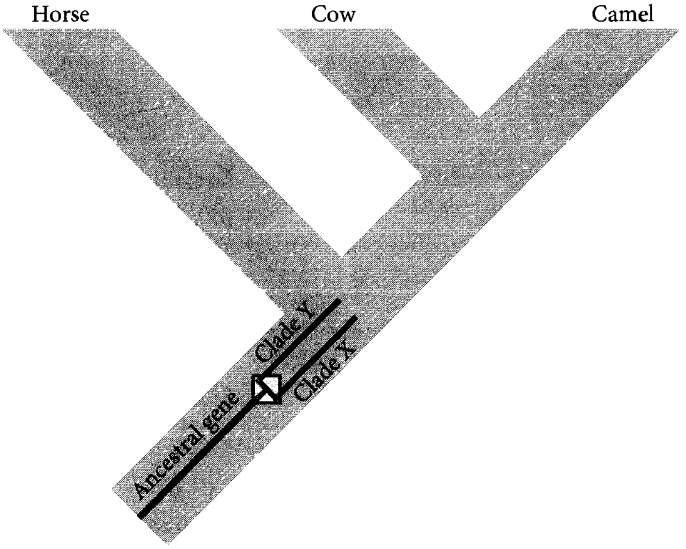


FIGURE 6.12 The first step in reconciling the trees in Figure 6.11. The gene duplication (marked with a square) giving rise to gene clades X and Y predated the first split in the population tree.

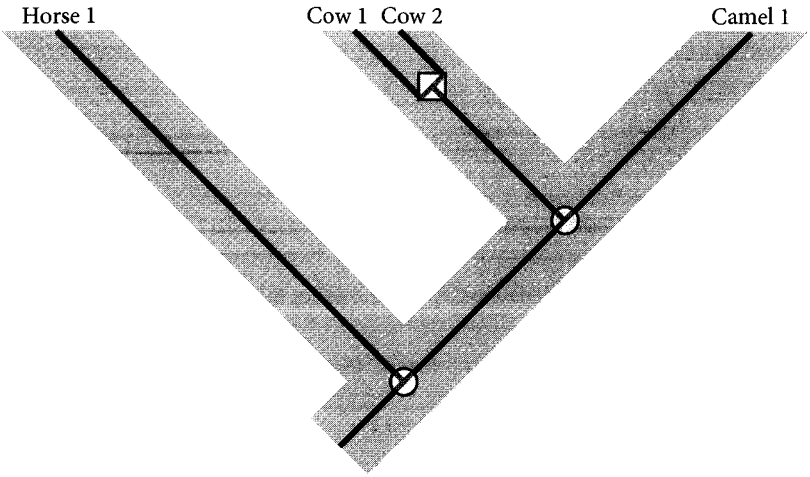


FIGURE 6.13 Reconciliation of clade X of the gene tree with the species tree. This requires one gene duplication event (marked with a square). The other nodes in the gene tree (marked with circles) correspond to population lineage-splitting events.

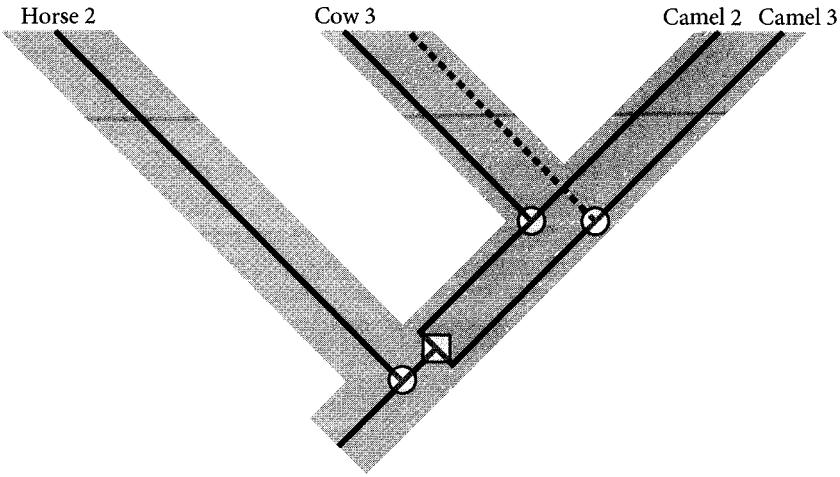


FIGURE 6.14 Reconciliation of clade Y with the species trees. This requires one gene duplication event (marked with a square) and one gene lineage extinction event (marked with a dashed line). All other nodes in the gene tree (marked with circles) correspond to population lineage-splitting events.

parsimonious that this duplication happened on the cow's terminal branch of the species tree, because otherwise the duplicates should be present in camels. We can now fit clade X into the species tree (Figure 6.13). This required one gene duplication event and no gene extinction events.

Within clade Y, horse 2 is sister to all of the other gene copies, suggesting that all the extra copies in cows and camels arose via gene duplication after the horse lineage diverged from the cow + camel clade. The latter clade contains two gene copies in camel and one in cow. Because camel 2 is more closely related to cow 3 than to camel 3, we can conclude that the duplication that gave rise to camel 2 and camel 3 must have predated the cow-camel split. Given this inference, the fact that there is no cow gene sister to camel 3 suggests that there has been a gene deletion in the cow lineage. Thus, we can explain clade Y by invoking one gene duplication event at the base of the cow + camel clade and a gene loss in the terminal branch leading to the cow. This hypothesis is shown in Figure 6.14.

Overall, we can reconcile this gene tree and this species tree by invoking three gene duplication events and one gene deletion event. This example should serve to show how an ability to interpret trees allows one to make inferences about how gene families have diversified over their evolutionary history.

POPULATION RETICULATION

The preceding sections have only considered population and species trees that are strictly treelike, with population lineages that split but never reconnect. In reality, evolution is not so tidy. Instead, population lineages sometimes adopt a netlike, or ***reticulate***, pattern.

Three phenomena cause population reticulation: ***lateral gene transfer*** (sometimes called ***horizontal gene transfer***), ***introgression***, and ***lineage fusion***. These three processes have different implications for the distribution of gene trees in the genome.

Lateral gene transfer (LGT) occurs when a small piece of the genome is transferred between organisms by a process other than traditional sexual reproduction. It is quite common in prokaryotic organisms but apparently relatively rare in eukaryotes such as animals and land plants. Figure 6.15 shows an example in which a piece of DNA (represented by the black arrow) is transferred from an individual in population lineage B to population lineage D. One way this

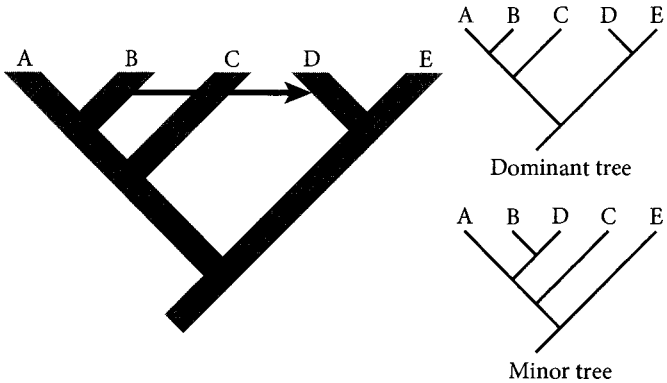


FIGURE 6.15 Example of the effect of lateral gene transfer on gene trees. Lateral gene transfer (represented by the dark arrow) results in the transferred genes having a different history (the minor tree) than the rest of the genome, which is expected to have the dominant tree.

can happen is if a virus infects one organism, acquires a segment of DNA from its host, and then transfers that piece of DNA to a second, perhaps distantly related host. If the individual in the recipient lineage (lineage D in Figure 6.15) passes on that new piece of gene, it can eventually become fixed in the population. In this case, all the individuals in the D lineage will have a small piece of their genome with a gene tree (the ***minor tree*** in Figure 6.15) that is different from the gene tree that applies to most of the genome—the ***dominant history***.

Introgression occurs when organisms from distinct population lineages come into contact and reproduce sexually, producing hybrid offspring that then breed with members of one or the other parental population. Figure 6.16 illustrates this process. You can imagine an organism from population C dispersing into population D and then crossing with D individuals. Just after the introgression, some D individuals would carry alleles that originated from the C population and would therefore have a distinct gene tree from the other organisms in the population. Assuming that the populations then remained separate, lineage sorting would occur within each population lineage. For most genes, the rare alleles that were introduced from population C would be lost by gene lineage extinction. However, for some parts of the genome, the "foreign" C gene will be the allele that goes to fixation. The gene trees for such regions will contradict those from the rest of the genome. In our hypothetical example, most of the genes in the genome will show the dominant history (matching the

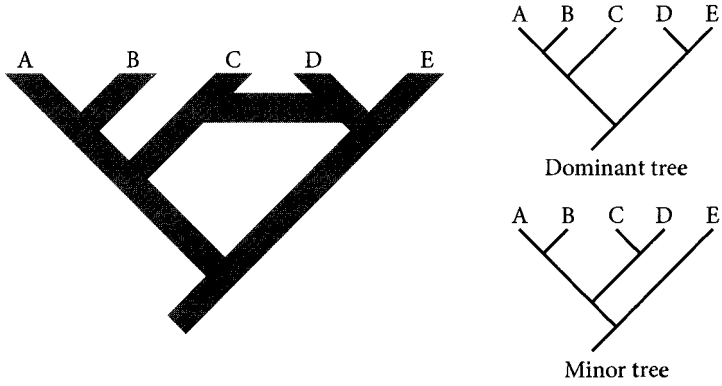


FIGURE 6.16 Example of the effect of introgression on gene trees. The dominant tree follows the population tree although introgression of genes from C into D can result in some genes having a minor history that places C sister to D.

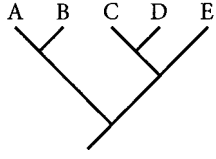


FIGURE 6.17 Directionality and introgression. If gene flow occurred by hybridization of C and D, followed by backcrossing of hybrids with members of D, then the minor tree will differ from that shown in Figure 6.16 in that the (C, D) clade will occupy the position formerly occupied by D.

population or species tree), whereas the genes that were introgressed from C to D would show a different history (the minor tree) with C and D as each other's closest relatives.

The fact that individuals from C and D came into contact and were able to interbreed makes it likely that there was also introgression in the reverse direction. This would occur if one or a few D individuals entered a population of species C and interbred. In this case we would expect a few D genes to become fixed in species C, resulting in a second minor tree, shown in Figure 6.17.

An extreme case of introgression is called lineage fusion. This entails the complete merging of two formerly distinct population lineages into a single descendant lineage. This phenomenon is sometimes called ***hybrid speciation***,

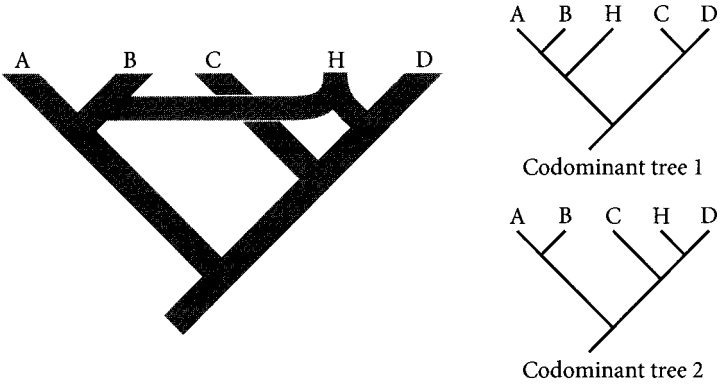


FIGURE 6.18 Example of the effect of lineage fusion on gene trees. Lineage fusion is expected to result in codominant histories, that is, conflicting trees with more or less equal frequencies in the genome. In this example, H is the hybrid lineage.

but given ambiguity over the terms "species" and "speciation," we prefer to call the phenomenon lineage fusion. When the two parent lineages contribute roughly equal numbers of individuals to the new hybrid lineage, we can expect that about half of the genes in the hybrids will group with one parent and the other half with the other parent. Thus, we would expect to see two ***codominant trees*** as shown in Figure 6.18.

TREE THINKING GIVEN RETICULATION

Does the existence of reticulation in population trees undermine tree thinking and the discipline of phylogenetics? The answer is no, for two main reasons.

First, although the phenomena described above do occur, many groups of organisms have predominantly treelike population histories. Introgression and lineage fusion are phenomena that require sexual reproduction: mating and the production of fertile offspring. In sexually reproducing organisms, the capacity to mate successfully is influenced by numerous biochemical, physiological, behavioral, and physical traits. Evolutionary divergence in any of these traits has the potential to cause reproductive isolation, that is, the prevention of successful reproduction. The longer that two lineages evolve separately, the more likely it is that traits will evolve in one or both lineages that result in reproduc-

tive isolation. As a result, successful mating is often possible only between individuals that are recently diverged. For this reason, introgression and lineage fusion tend to be local affairs, affecting only closely related lineages. Introgression and lineage fusion might lead to a non-treelike history among, say, species of mice, but they would be unlikely to occur between, say, mice and hamsters.

In the case of lateral gene transfer (LGT), there are no constraints on the phylogenetic "breadth" of reticulation: a gene can move as readily from a bacterium into a plant as from a crocodile into an alligator. In many cases, however, such inserts from distant taxa are likely to be lost via lineage sorting, especially if the gene offers no benefit to the recipient. Also, LGT involves rather minute pieces of DNA and, thus, impacts only a tiny minority of the genome. Only in groups in which LGT is rampant and the inserts are large relative to the rest of the genome, as may be the case for some bacteria and archaea, is the tree model truly undermined. Although we have much to learn about the extent of LGT, the fact that we often recover the same or similar trees from different genes for the same sets of organisms suggests that LGT has not obscured our capacity to detect the dominant history of population lineages in many kinds of organisms.

A final point to stress is that, even when the population history is reticulate, gene trees are not. Recall that within the fully reticulate genealogy of a sexual population, recombinational genes always have strictly treelike histories (e.g., Figure 6.4). Thus, tree thinking is valid and useful even when population lineages do not form trees. Indeed, it is by studying the discordance among gene trees that biologists have the means to learn about the shape of population histories. Thus, rather than undermining the phylogenetic paradigm, reticulate evolution provides one more reason for understanding trees.

THE CONCEPT OF SPECIES

Having introduced the concept of a population tree, it is appropriate to touch on one of the most controversial topics in evolutionary biology—the definition of "species." The concept of species has long been a source of controversy among biologists. One might have thought that the adoption of a phylogenetic outlook in evolutionary biology would have resulted in some degree of unanimity, but this has not happened. Evolutionary biologists still struggle to clarify the relationship of "species" to different kinds of genealogies—organismal pedigrees, population trees, and gene trees.

The core of the so-called species problem is that there is no single definition of species that achieves all that has been historically expected of species. In an ideal world, we could devise a definition of species such that they are the entities that participate in evolution (predicting the future) and are also the products of evolution (reflecting the past). Different popular definitions of species succeed in making them either players in evolution, products of evolution, or neither—but never both. Here we briefly summarize some of the alternative approaches to thinking about species and how these different species concepts relate to trees.

The first way of thinking about species is as groups of organisms sharing a particular trait. We can call species concepts in this vein ***trait-based species concepts***. This is a broad umbrella. Individual species definitions within this umbrella differ in whether the attribute uniting members of a species is a single trait or a set of traits, or whether certain kinds of features, for example, ecological specialization or reproductive capacity, are given priority.

While trait-based species concepts have the practical advantage that species can generally be distinguished by visible traits, they do not align tidily with phylogenetic history. The reason is that organisms can share similar traits, not because they are closely related, but because they have retained the same ancestral character state. For example, Figure 6.19 depicts a population tree on which key traits have been mapped. If we define species as groups of organisms that have the same state of these key traits, then three species are recognized. Species B and C are defined by derived character states, whereas species A is composed

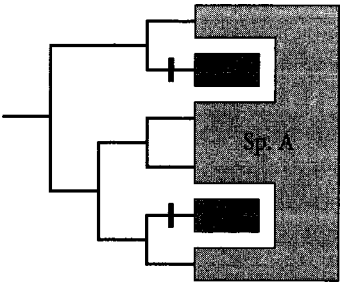


FIGURE 6.19 Delimiting species based on sharing particular traits can result in individuals of the same species not being each other's closest relatives. Trait evolution is denoted with a black bar. Members of species A are united by a shared lack of the derived traits and do not form an exclusive group.

of lineages that have retained the ancestral state. In this case, species A does not correspond to a monophyletic (exclusive) group: some members of species A are more closely related to members of species B or C than to other members of species A. This phenomenon will apply regardless of whether the trait in question is morphological, behavioral, physiological, molecular, ecological, or geographical.

Whereas some trait-based species concepts focus on individual traits, the ***phenetic species concept*** focuses on overall similarity (the variation seen in all traits). This concept runs into the same problem as other trait-based concepts when the rate of evolution differs among lineages, as will often be the case. Figure 6.20 shows a phylogram in which the branches are drawn proportional to the actual amount of evolution occurring on each lineage. Whereas most of the tree experienced the same slow rate of evolution, two terminal lineages have undergone much more rapid evolution (approximately ten times as fast). The members of all three phenetic species are similar to each other and dissimilar to members of the other two species. However, in the case of species A, the high degree of similarity within the species reflects the shared *lack* of change rather than true evolutionary kinship. Again, some members of species A are more closely related to species B (or C) than to other members of species A. So, regardless of whether a trait-based concept is based on individual traits or overall similarity, it fails to yield a concept of species that consistently aligns with evolutionary history.

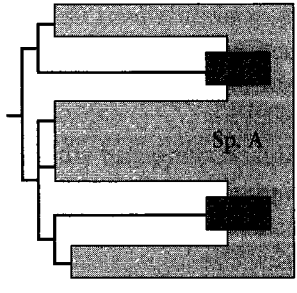


FIGURE 6.20 **Delimiting species based on overall similarity can result in individuals of the same species not being each other's closest relatives.** A species concept based on overall similarity would distinguish species B and C from the rest of the tips (species A) due to the fact that species B and C experienced rapid evolution along their terminal branches. Members of species A are united by a shared lack of rapid evolution and do not form an exclusive group.

Another trait-based species concept that warrants special attention is the ***biological species concept***, which is popular among many zoologists and widely used in textbooks. This defines a species as a set of populations that have the potential to interbreed with one another, but have reproductive features that prevent them from interbreeding with members of other biological species. The biological species concept is centered on one particular trait, reproductive compatibility, and runs into the same problem as other trait-based concepts: not mapping onto the tree of evolutionary relationships.

Within populations we expect organisms to share the same reproductive system and to be reproductively compatible. However, over time, evolutionary changes to the reproductive system occurring in a population lineage can cause members of that lineage to lose the ability to interbreed with populations that have retained the ancestral reproductive system. Applying the biological species concept, the population with a derived reproductive system will be defined as a new species. At the same time, all of the populations that share the ancestral reproductive system would be considered conspecific (i.e., members of the same species) because they still can interbreed, despite the fact that they need not comprise a clade. So, as with other trait-based concepts (Figure 6.19), the biological species concept can often place groups of contemporary individuals into the same species even when they are not each other's closest relatives, thereby misrepresenting evolutionary history.

The alternative strategy to using traits as a basis for defining species is to define species as groups of organisms that are united by a shared history. This is the strategy of ***phylogenetic species concepts***. Under this approach, species are composed of organisms or populations that are exclusive/monophyletic. In contrast to the trait-based concepts, phylogenetic approaches to species definitions view species as products of evolution rather than active players in evolution. While there is variation and controversy as to how exactly to apply the concepts of monophyly/exclusivity and whether to focus on gene trees or organismal pedigrees, most phylogenetic species concepts share the objective of viewing species as clades that are directly comparable to, and tidily nested relative to, taxa at other ranks. The logic is that, as with other taxa (see Merging and Pruning in Chapter 3), every organism sampled from a phylogenetic species should occupy the same place on the tree of life as all other organisms sampled from that species.

Although the phylogenetic species concept may appeal because it resonates with tree thinking, it does not fulfill all of the desirable properties of a species

concept. Most notably, the *ranking* of a phylogenetic species is not strictly objective. The criterion used to group organisms into species is the same criterion that we use to delimit taxa at other ranks. If organisms are grouped into subspecies or genera using the same criteria that are used to assign organisms to species, then the species rank is not special. Although there are a number of principles that can be used to guide the recognition of a taxon at the species rank (e.g., species are understood to show relatively little morphological or ecological variation and usually have a homogenous reproductive system, especially in a single locality), the distinction between a species and a genus is quantitative, not qualitative. If you view the species rank as special, for example, if you think that species participate in evolution in a qualitatively different way than subspecies, genera, and so on, then you would likely favor trait-based over phylogenetic species concepts.

A final important point to stress is, whatever you think the term "species" means, the actual groups of organisms that are assigned species names may not meet this criterion. Even if you agree that species *ought* to be understood as being monophyletic groups, this should not be mistaken for the view that actual named species, *Canis familiaris, Xenopus laevis, or Magnolia grandiflora*, necessarily correspond to monophyletic groups of organisms. No doubt there are some or many cases where they do. Evidence suggests that *Homo sapiens* fits this criterion. However, there are also many cases where species names have been found not to align with monophyletic groups. For example, research has suggested that brown bears do not correspond to a monophyletic group because some brown bears (those in certain islands close to Alaska) are more closely related to polar bears than to other brown bears.

We should not be surprised that many named species are not monophyletic. Most taxonomists who describe species have not adopted a fully phylogenetic perspective and describe new species whenever they find groups of organisms with distinctive morphological features, or ***diagnostic traits***. This approach is often motivated by practicality as opposed to the search for monophyletic groups or the conscious application of a trait-based species concept. Still, many groups that are delimited by this approach will actually be monophyletic.

Furthermore, even if taxonomists consciously subscribe to the view that species names should refer to monophyletic taxa, they would still make mistakes. Naming a phylogenetic species (or any other taxon) is the same thing as proposing the hypothesis that the group of organisms is monophyletic. For the vast majority of named species, the amount of evidence available to assess monophyly is minimal. Often, our evidence for the monophyly of described species

depends on a handful of museum or herbarium specimens with scant geographic information. Nonetheless, with the increasing ease of collecting genetic data, it is becoming much easier to directly test the hypothesis of monophyly for many traditionally recognized species.

FURTHER READING

Coalescence and incomplete lineage sorting: Maddison 1997; Chapter 1 of Wakeley 2009

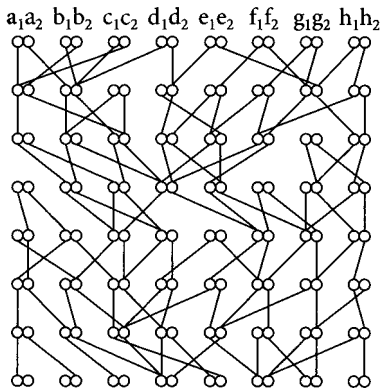
Gene duplication: Fitch 1970; Page and Holmes 1998; Graur and Li 2000; Boussau and Daubin 2010

Population reticulation: Doolittle 1999; Baum 2007; Boussau and Daubin 2010

Phylogenies and species concepts: Mishler and Donoghue 1982; Donoghue 1985; Ereshefsky 1989, 2011; Baum and Shaw 1995; Mayden 1999; de Queiroz 2005; Baum 2009; Velasco 2009

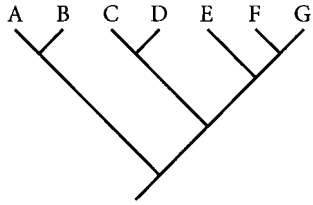
CHAPTER 6 QUIZ

The following questions refer to the pedigree. Each diploid individual is represented by a pair of circles that represent the two alleles of the gene under consideration.

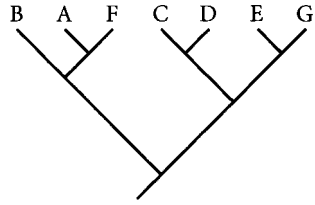


1. What do the lines in the figure represent?
2. Is allele b1 more closely related to b2 or a2?
3. What is the gene tree for alleles b1, b2, c2, and f1?
4. What is the gene tree for alleles a2, b2, c2, and g1?
5. All of the following alleles coalesce within the series of generations shown, except one. Which is the exception: a1, a2, b1, b2, f1, f2, h1, or h2?
6. How many alleles from the lowermost generation have at least one descendant in the uppermost generation?

7. Tree 1 is true for 99.9% of the genome and tree 2 for 0.1% of the genome. What is the most plausible explanation?
   a. Introgression between A and F
   b. Introgression between B and an ancestor of A and F
   c. F is a hybrid between A and G
   d. Lateral gene transfer from A to F
   e. Lateral gene transfer from F to A

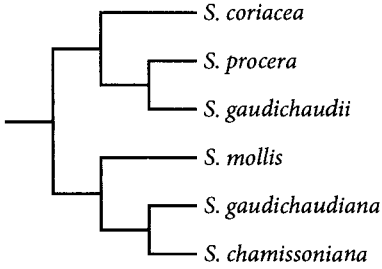


Tree 1

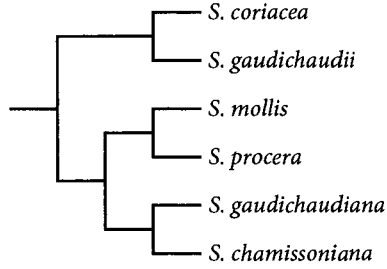


Tree 2

8. Howarth and Baum (2005) estimated the phylogenetic relationships of six species of Hawaiian shrubs in the genus *Scaevola*. They inferred that about half the genes had tree 1 and about half had tree 2. What is the most likely explanation?
   a. Introgression between *S. procera* and *S. mollis*.
   b. Introgression between *S. procera* and *S. gaudichaudii*.
   c. *S. procera* is a hybrid between *S. mollis* and *S. gaudichaudii*.
   d. There was lateral gene transfer from *S. procera* into an extinct ancestor of *S. gaudichaudiana* and *S. chamissoniana*.
   e. There was lateral gene transfer from *S. mollis* into *S. procera*.

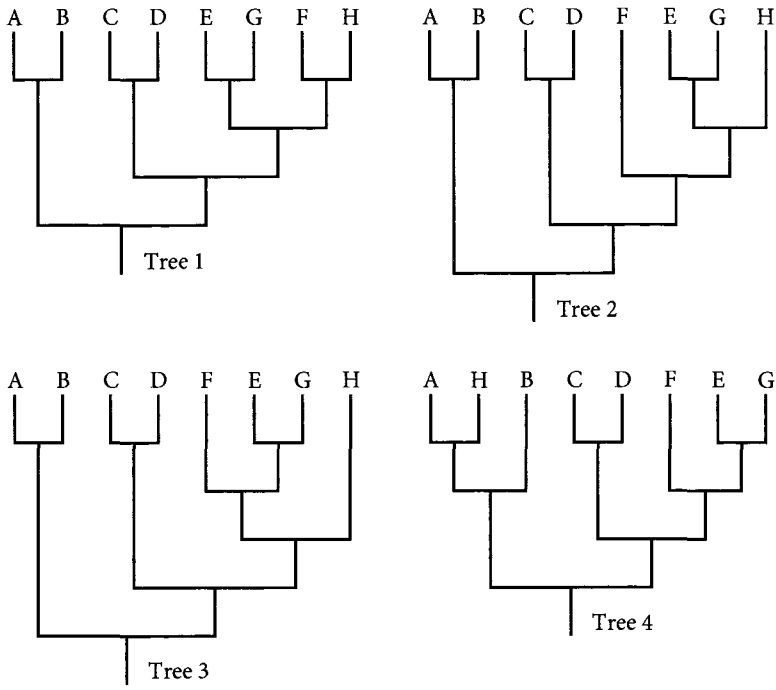


Tree 1

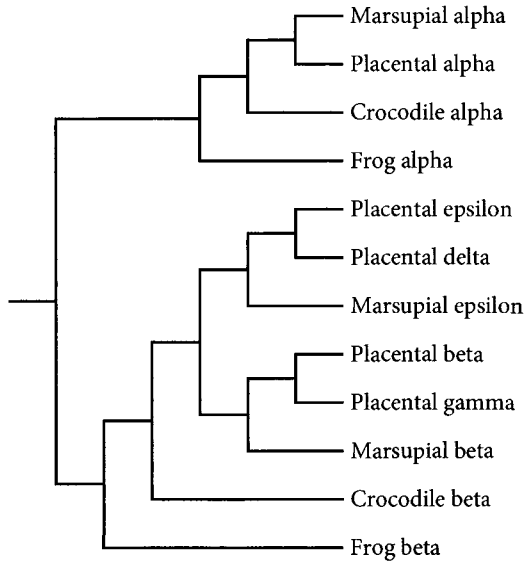


Tree 2

9. For a set of 8 organisms, 33% of the genome has tracked each of trees 1, 2, and 3 and 1% has tracked tree 4.
   a. For what proportion of the genome is clade (EFG) true?
   b. For what proportion of the genome is clade (EFGH) true?
   c. For what proportion of the genome is clade (FH) true?



Questions 10–12. The figure is the gene tree for several hemoglobin genes from marsupials, placental mammals, crocodiles, and frogs. Based on this tree, answer the following questions.

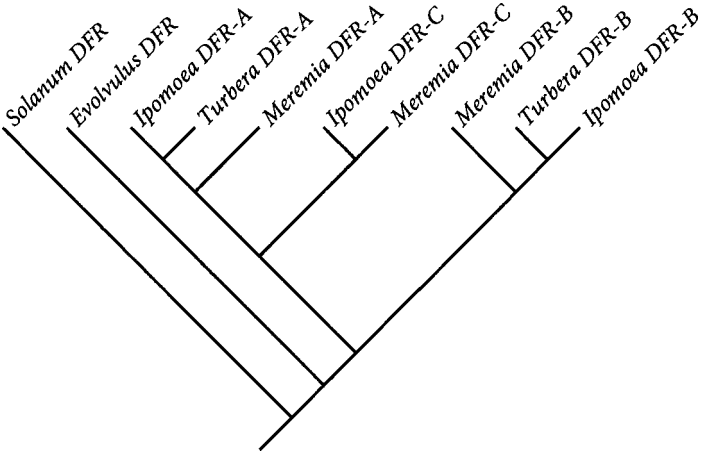


10. Which of the following genes is an ortholog of the marsupial hemoglobin alpha?
    a. Marsupial beta b. Placental gamma c. Frog alpha
    d. Crocodile beta e. Marsupial epsilon

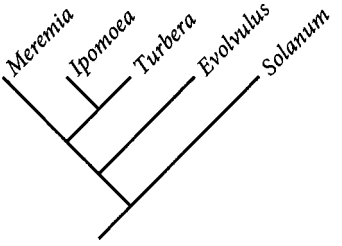
11. Which of the following genes is an ortholog of the crocodile hemoglobin beta?
    a. Crocodile alpha b. Frog beta c. Placental beta
    d. Marsupial epsilon e. Placental gamma

12. How many gene duplication events separate the frog hemoglobin beta from placental hemoglobin beta?
    a. 0 b. 1 c. 2 d. 4 e. 5

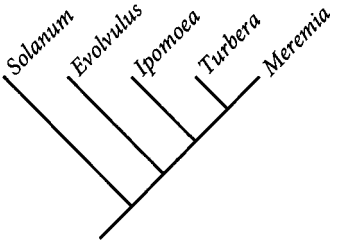
Questions 13–14. Assume that the inferred gene tree for members of the DFR gene family from five species in the morning glory and tomato families is correct (based on Des Marais and Rausher 2008).



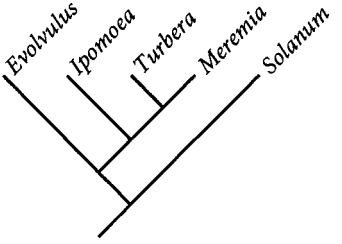
13. Which species tree is most plausible given this gene tree?



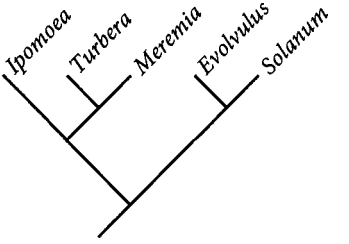
a



b



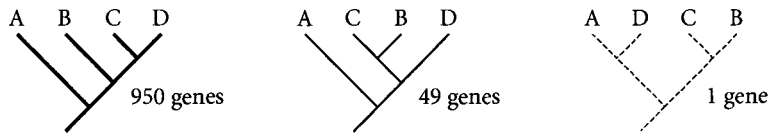
c



d

14. Which of the following is the most plausible history of gene duplication and gene loss given parsimony?
    a. Three gene duplications occurred on the lineages leading to *Ipomoea* and *Meremia*, and two occurred on the lineage leading to *Turbera*.
    b. Two rounds of gene duplication occurred before the entire group diversified, with gene copies then lost (or overlooked) in *Solanum*, *Evolvulus*, and *Turbera*.
    c. Two rounds of gene duplication occurred in the lineage leading to *Ipomoea*, *Meremia*, and *Turbera*, but one gene was lost (or overlooked) in *Turbera*.
    d. One gene duplication occurred in the last common ancestor of *Ipomoea*, *Meremia*, and *Turbera*, and a second gene duplication occurred in the last common ancestor of *Ipomoea* and *Meremia*.
    e. One gene duplication occurred in the last common ancestor of *Ipomoea* and *Meremia* with subsequent introgression of one copy into *Turbera*.

15. Imagine we have sequenced 1000 genes for 4 taxa (A, B, C, D) and find three topologies for the indicated number of genes.
    a. Draw the population (species) tree that minimizes the number of deep coalescence events.
    b. Fit the three gene trees within this population tree, assuming that all discordance is due to incomplete lineage sorting.
    c. Given the proportion of genes showing each of the three histories, and assuming that all gene trees are accurate, is incomplete lineage sorting likely to have generated these differences across genes? Why or why not?



16. Assume that a certain major clade of animals is characterized by a strictly divergent population tree. If the biological species concept is applied to this group, would we expect all species to correspond to clades on the population tree? Why or why not?

CHAPTER SEVEN

Phylogenetic Inference with Parsimony

Given the importance of phylogenetic trees in modern biology, it is important to know enough about the reconstruction of phylogenetic trees, ***phylogenetic inference***, to understand why we might (or might not) wish to accept trees generated by evolutionary biologists. Additionally, phylogenetic inference serves as an excellent example of the general principles that allow scientists to elucidate events that happened in the past. As a result, even if you are never likely to do phylogenetic research, it is worth knowing something about phylogenetic analysis so that you can appreciate the rigor of historical sciences such as evolutionary biology, geology, paleoclimatology, and cosmology.

Nowadays there are several alternative methods for phylogenetic inference, most of which proceed via the same basic steps: constructing a data matrix, identifying trees that are most compatible with the data matrix, and then conducting statistical analyses to evaluate how confident we should be in our phylogenetic conclusions. In this chapter we describe the first two steps in this process, focusing on the method of maximum parsimony and its historical predecessor, Hennigian inference. Parsimony is just one of a variety of methods for phylogenetic inference. It provides a useful starting point for understanding how phylogenetic trees are estimated and can serve as a foundation for the introduction of model-based methods (Chapter 8).

A BIOLOGICAL EXAMPLE: CARNIVORA

To provide a context for the discussion of methods of phylogenetic inference, we will use a simplified biological example, a study of the Carnivora. Carnivora

is a group of mostly meat-eating mammals, including dogs, cats, bears, weasels, mongooses, skunks, and seals. While these animals differ greatly in their external appearance and ecology, they share several skeletal features. For example, almost all species have enlarged side teeth, carnassials, which may be used for shearing meat (Figure 7.1), six incisors, and two well-developed canines in each jaw. Based on these and other traits, it has long been accepted that the Carnivora is a monophyletic group, a clade. But what are the relationships within Carnivora?

Before embarking on a study of Carnivoran phylogeny, we need to decide which species to select for our study and which traits to score. With around 250 carnivoran species, we cannot easily examine all living forms. How many and which species to include in a study is governed somewhat by the specific questions we wish to answer. Let us say that our pressing concern is to find out if the carnivores are divided into two monophyletic subgroups, the aquatic pinnipeds (seals, sea lions, and walruses) and the terrestrial fissipeds (all others). This is a long-standing hypothesis that predates the development of formal phylogenetic methods. To answer this question, we would have to sample representative species from most of the carnivoran families (dogs, cats, seals, sea lions, etc.). While a research scientist would likely include multiple representatives for each family (to test family monophyly and minimize the chances of artifactual results), we will use just one species from each of ten families (three pinnipeds and seven fissipeds) to simplify the example.

Now that we have chosen which taxa to include, we must decide which characters to use. Any trait that varies among tips and is thought to show some

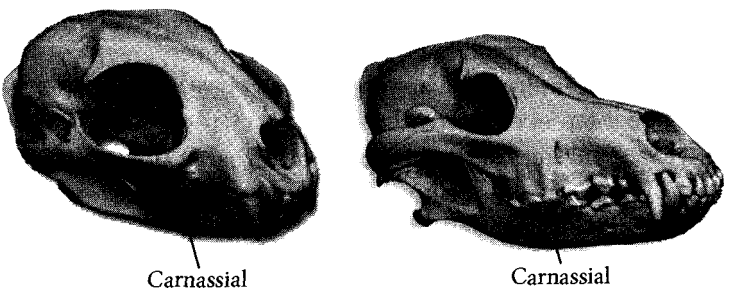


FIGURE 7.1 Bobcat and Mexican grey wolf skulls. Upper carnassial teeth are indicated. Images used with permission from www.SkullsUnlimited.com.

degree of heritability has the potential to provide phylogenetic information. Appendix 1 summarizes the main classes of data that may be used for phylogenetic analysis. Until the early 1990s, phylogenetic analyses were usually based on morphological traits. With the advent of modern molecular methods, almost all phylogenetic studies now employ DNA sequence data. However, it will be instructive to begin with a consideration of morphological characters. We will do this by extracting morphological data from a published study by Wyss and Flynn (1993).

There is one final consideration before we can begin to collect data: we must choose one or more ***outgroup*** taxa. Outgroups serve as a point of comparison with the ***ingroup*** (here, the carnivorans), allowing us to root our trees and determine the direction (polarity) of character change. Any taxon that is not a member of Carnivora could theoretically serve as a valid outgroup. However, the best outgroups are reasonably closely related to the ingroup so that traits are more easily compared between the ingroup and outgroup. For our analysis of morphological data we will use as an outgroup the creodonts, an extinct group of mammals. The reason that they can serve as an outgroup is not because they are extinct. Rather, it is because they fall outside of the Carnivora clade as indicated by the fact that they lack certain shared derived characteristics or synapomorphies of Carnivora, such as the bony auditory bulla that encloses the inner ear. A living group that is known to be outside the Carnivora clade would do just as well provided that it had characters that could readily be compared to those found in Carnivora.

Now we can proceed to score the ingroup and outgroup taxa for the morphological traits selected. Appendix 2 provides more information on the complexities of building a data matrix based on morphological data. The species are scored for each trait by observing an individual or a few individuals from that species and recording the form of that trait, its character state. The 12 characters and the states for each character are given in Table 7.1. We have assigned 0 to all of the character states present in the outgroup. However, no significance should be attached to this convention: choosing different labels for the states would not affect the results.

Moving through the tips, we record the character state for each character for each species to build a ***character state matrix***. For example, we observe that, for trait 6 (the tail), creodonts have the "elongated" state. We have chosen to represent this state with a 0, so creodonts are given a score of 0 for trait 6 in the

TABLE 7.1 Characters and character states for an analysis of carnivorans, with numerical representation (as used in Table 7.2) provided in parentheses

No.	Character	States
1	Complexity of the mucus-coated surfaces in the nose (maxilloturbinals)	Minimally branched (olfactory surfaces in nasal passage) (0); highly branching (olfactory surfaces excluded from the nasal passage) (1)
2	Bony spur by the auditory bulla (paroccipital process)	Straight and projecting (0); cupped around auditory bulla (1)
3	Number of lower incisors	2 (0); 3 (1)
4	Upper molar 1	Present (0); absent (1)
5	Baculum (bone within the penis)	Present (0); absent (1)
6	Tail	Elongated (0); short (1)
7	Hallux (5th digit, or dewclaw, on hind leg)	Prominent (0); reduced or absent (1)
8	Claws	Nonretractable (0); retractable (1)
9	Prostate gland	Small and simple (0); large and bilobed (1)
10	Kidney structure	Simple (0); conglomerate (1)
11	External ear (pinna)	Present (0); absent (1)
12	Testis position	Scrotal (0); abdominal (1)

matrix. Likewise, dogs, cats, and many other carnivorans have long tails so they also get state 0, whereas bears, seals, sea lions, and walruses have short tails and are assigned state 1.

The complete character state matrix for the 12 characters for 10 ingroups and 1 outgroup is given in Table 7.2. Notice that some taxa may be scored as unknown for certain characters (conventionally represented with '?'). This could be because we are ignorant as to the proper scoring (e.g., soft tissues in a fossil) or because it is impossible to score (e.g., toe number in snakes). While not present in this matrix, a taxon whose members may express different

TABLE 7.2 Morphological character state matrix for carnivorans, with creodont included as the outgroup

	Character state scoring											
Taxon	1	2	3	4	5	6	7	8	9	10	11	12
Creodont	0	0	0	0	0	0	0	0	?	?	?	?
Cat	0	1	0	1	0	0	1	1	1	0	0	0
Hyena	0	1	0	1	0	0	1	0	1	0	0	0
Civet	0	1	0	0	0	0	0	0	1	0	0	0
Dog	1	0	0	0	1	0	0	0	0	0	0	0
Raccoon	1	0	0	0	1	0	0	0	0	0	0	0
Bear	1	0	0	0	1	1	0	0	0	1	0	0
Otter	1	0	0	0	1	0	0	0	0	1	0	0
Seal	1	0	1	0	1	1	0	0	0	1	1	1
Walrus	1	0	1	0	1	1	0	0	0	1	1	1
Sea lion	1	0	1	0	1	1	0	0	0	1	0	0

character states can be scored as ***polymorphic*** by listing multiple states within a cell.

As you might imagine, it can be difficult to find a large number of morphological traits that show appropriate levels of variation for reconstructing a phylogeny. In comparison, it has become quite easy to obtain large amounts of DNA sequence data. Table 7.3 shows some DNA sequence data for the carnivorans. Because DNA is unavailable for creodonts, an alternative outgroup, a mole, has been substituted. Like the creodonts, we can be sure that moles are outside the Carnivora clade.

Whereas for morphological data the character states were coded as 0's and 1's, the states of DNA are the 4 bases (A, C, G, and T). You may also observe ***gaps*** (marked with a hyphen) in the DNA matrix. These gaps arise when bases are inserted or deleted during the course of evolution. The process of ***sequence***

TABLE 7.3 The states for 15 consecutive positions in the transthyretin 2 gene, with mole included as the outgroup

	Positions in DNA sequence														
Taxon	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
Mole	G	T	T	A	A	-	C	T	T	C	T	C	A	C	T
Cat	G	T	T	G	A	-	C	C	T	C	T	T	A	C	T
Hyena	G	T	T	G	A	-	C	C	T	C	T	C	A	C	T
Civet	G	T	T	G	A	-	C	C	T	C	T	C	A	C	T
Dog	G	T	T	A	A	G	C	A	T	C	T	G	C	C	T
Raccoon	G	T	T	A	A	G	G	G	T	C	T	G	C	C	T
Bear	C	T	T	A	A	G	T	G	T	C	T	G	C	C	T
Otter	G	T	T	A	A	G	G	G	T	C	T	G	C	C	T
Seal	G	T	A	A	A	G	C	G	T	C	T	G	C	C	T
Walrus	G	T	A	A	A	G	C	G	T	C	T	G	C	C	T
Sea lion	G	T	A	A	A	G	C	G	T	C	T	G	C	C	T

alignment is concerned with establishing the correct position of gaps so that homologous sequence positions are aligned above one another in the data matrix. Sequence alignment will be discussed more fully later in this chapter.

HENNIGIAN INFERENCE

In the middle part of the twentieth century, the German entomologist Willi Hennig and colleagues developed the first formal method for phylogeny reconstruction. This was described in Hennig's 1966 book, *Phylogenetic Systematics*, which is generally credited with launching the modern field of phylogenetics. While Hennigian inference (or Hennigian "argumentation") is no longer used, we believe it is worth knowing about. The method played an important role in the historical development of phylogenetics. It is simple to understand and illustrates the general point that the distribution of trait variation among

taxa contains information about evolutionary relationships. Moreover, by understanding the problems with Hennigian inference, and the reasons why it has been replaced by other methods, one can gain a clearer appreciation of the distinction between algorithmic and optimality methods for phylogeny reconstruction.

Hennigian inference makes two major assumptions: (1) There is a strictly treelike phylogenetic history, and (2) there is no homoplasy—each character evolved from an ancestral (plesiomorphic) to a derived (apomorphic) state once, without subsequent reversal. In other words, Hennigian methods require that there be no back mutations and no independent forward mutations. Given these assumptions, a set of tips sharing an apomorphic state must be a clade.

Let us apply this principle to the morphological data for Carnivora. The first problem is to determine which character states are ancestral and which are derived. For example, did dewclaws evolve within the group, or were dewclaws lost? While many methods have been proposed for determining character polarity (Chapter 4), the most widely used is the outgroup method. If a state is variable in the ingroup, the state that occurs in the outgroup is the ancestral state for the ingroup. If no character states are shared between the ingroup and the outgroup, the ancestral state for the ingroup cannot be determined. In this case the outgroup creodonts have dewclaws (character 7, Tables 7.1 and 7.2), so "dewclaws present" is the plesiomorphic or ancestral character state.

You may notice that creodonts are scored as uncertain ('?') for the soft tissue traits: kidneys, prostate gland, ears, and testes. By examining additional living outgroups, we can determine that the ancestral state for these four characters was probably also state 0. A modified data matrix with a hypothetical outgroup having all ancestral states is shown in Table 7.4.

The Hennigian method of phylogenetic inference involves identifying sets of taxa that share a derived character state and inferring that they form a clade. For example, character 11 supports a seal + walrus clade, and character 4 supports a cat + hyena clade. Applying this method we can now draw a tree that contains all of the clades suggested by the 12 morphological characters (Figure 7.2). To make it easier to interpret, we have indicated the numbers of the characters that support each clade.

Prior to Hennig, scientists had lacked well-defined protocols for phylogeny reconstruction. Hennigian inference provided a clear, objective method for using observed trait variation to reconstruct evolutionary history. As a result, this method revolutionized evolutionary biology and stimulated the emergence

TABLE 7.4 Morphological data matrix for carnivorans

	1	2	3	4	5	6	7	8	9	10	11	12
Outgroup	0	0	0	0	0	0	0	0	0	0	0	0
Cat	0	1	0	1	0	0	1	1	1	0	0	0
Hyena	0	1	0	1	0	0	1	0	1	0	0	0
Civet	0	1	0	0	0	0	0	0	1	0	0	0
Dog	1	0	0	0	1	0	0	0	0	0	0	0
Raccoon	1	0	0	0	1	0	0	0	0	0	0	0
Bear	1	0	0	0	1	1	0	0	0	1	0	0
Otter	1	0	0	0	1	0	0	0	0	1	0	0
Seal	1	0	1	0	1	1	0	0	0	1	1	1
Walrus	1	0	1	0	1	1	0	0	0	1	1	1
Sea lion	1	0	1	0	1	1	0	0	0	1	0	0

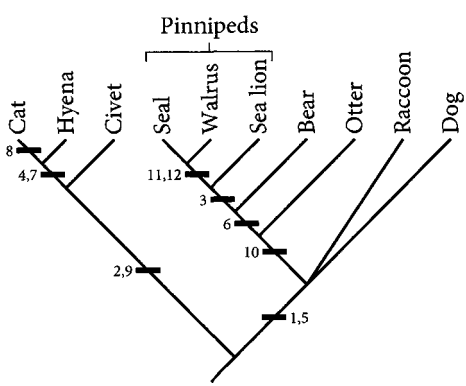


FIGURE 7.2 Phylogeny deduced from the data in Table 7.4 using Hennigian inference. Bars represent the origin of the derived character state of the character number(s) indicated.

of many other methods of phylogenetic inference. Although the Hennigian method described here was popular in the 1960s and 1970s, it is no longer in use. The core problem of Hennigian inference is that it makes unrealistic assumptions about trait evolution and provides no clear way of proceeding when those assumptions are not met.

The Hennigian method assumes that homoplasy is absent. But homoplasy does happen. Some characters secondarily evolve to closely resemble an ancestral state, and sometimes two indistinguishable traits evolve in parallel. Furthermore, even if evolution were strictly Hennigian, we should not expect all the traits, *as we score them*, to behave in a strictly Hennigian manner. We will sometimes make mistakes in character scoring, and we will sometimes make errors in the determination of character polarity.

If we examine additional carnivoran characters from the larger morphological matrix put together by Wyss and Flynn (1993), we can quickly find cases that serve to prove that Hennig's rules do not apply. Consider a thirteenth character: the presence/absence of lower premolar 1. This tooth occurs in the outgroup and all ingroups except cat, hyena, and otter. The absence of lower premolar 1 in cat, hyena, and otter suggests that these three form a clade. However, such a clade would require homoplasy in several other characters, for example, character 1 (branching of the turbinal bones). A tree that is fully consistent with character 13 would be inconsistent with character 1. Because there is no tree that is consistent with both characters 1 and 13, we know that Hennig's rules were broken.

If we allow the possibility of homoplasy, it becomes possible to reconcile any data matrix with any tree. But in that case Hennig's logic can no longer be used to deduce the true tree. Once the model is violated, the Hennigian deductive logic cannot tell us which tree is correct. Instead we need an ***optimality criterion***, a metric that can be used to decide, given some data, which trees are better and which trees are worse. Among the first optimality criterion proposed to replace Hennigian inference was ***maximum parsimony***.

THE MAXIMUM PARSIMONY CRITERION

Once we acknowledge that traits sometimes show homoplasy, one logical way to proceed is to allow that some homoplasy occurred, but to minimize the amount of homoplasy. This is an application of the principle of parsimony. We introduced this principle in Chapter 4 as a method for reconstructing the evolutionary history of a character, given a tree. Here we are using it in a slightly different way. In this context, the maximum parsimony criterion holds that the best estimate of phylogeny is that tree which explains all of the observed data by invoking the least homoplasy, which is to say, the fewest character state

changes. Referring back to the consistency index (Chapter 4), parsimony selects the tree that maximizes the average consistency index of the characters in the matrix. The simplest implementation of parsimony proceeds in three steps:

1. For a single tree, we consider each character in turn and determine the minimum number of character state changes, or ***steps,*** that are required to account for the distribution of states among tips (see Chapter 4).
2. We sum up the number of steps required by each character. The number of steps required to explain all of the characters' evolution is called the ***tree length.***
3. We repeat the preceding steps for all alternative trees and then identify the tree with the lowest tree length, which is the ***shortest*** or ***most parsimonious tree.***

Before applying parsimony to the Carnivora data set, let us consider a simple data matrix for four taxa (Table 7.5). If we assume that taxon O (the outgroup) is the sister taxon to the remaining species, that is, that taxa A–C form a clade, then three trees are possible (Figure 7.3).

We can start by considering tree 1 and seeing how we can explain each character in turn. If tree 1 were true, the simplest way to explain the first character is that all lineages began with state 0, but that a single change to state 1 occurred

TABLE 7.5 A simple morphological data matrix for four taxa

	1	2	3	4	5	6	7	8
O	0	0	0	0	0	0	0	0
A	0	1	0	0	0	1	1	0
B	1	1	0	1	1	1	1	1
C	0	0	1	1	0	0	0	0

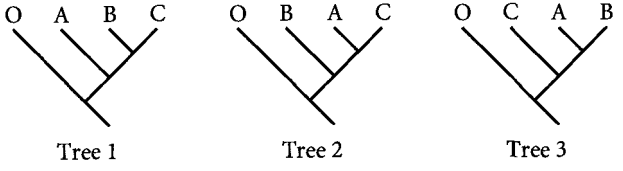


FIGURE 7.3 **The three possible rooted trees for four taxa.** Taxon O is the outgroup and taxa A, B, and C constitute the ingroup.

somewhere on the lineage leading to taxon B. Thus, we can explain this character with a single step: character 1 has length 1 on tree 1.

Character 2 is more difficult to map onto tree 1. There is no way to explain the distribution of states among the tips with only one change, but there are three ways to do it with two changes. These three equally parsimonious reconstructions are shown in Figure 7.4. The first scenario entails two independent transitions to state 1 (from state 0), the second entails two independent transitions to state 0 (from state 1), and the third entails one change to state 1 and one reversal back to state 0. When inferring a phylogeny, we do not need to know which of these reconstructions is correct. All that matters is that it takes a minimum of two changes to map character 2 onto tree 1: character 2 has length 2 on tree 1.

Using the same approach we can now map all eight characters onto tree 1. Characters 1, 3, 4, 5, and 8 each have only one most parsimonious reconstruction, whereas characters 2, 6, and 7 each have multiple, equally parsimonious reconstructions. For those characters, we have arbitrarily selected one of the most parsimonious mappings in Figure 7.5. In total, 11 steps are required to explain all the characters' evolution: tree 1 has a length of 11 for these data.

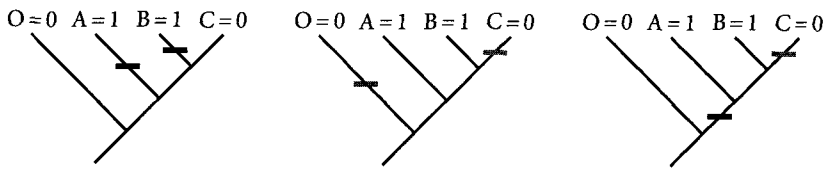


FIGURE 7.4 Alternative histories for character 1. Black bars: change from 0 to 1. Gray bars: change from 1 to 0.

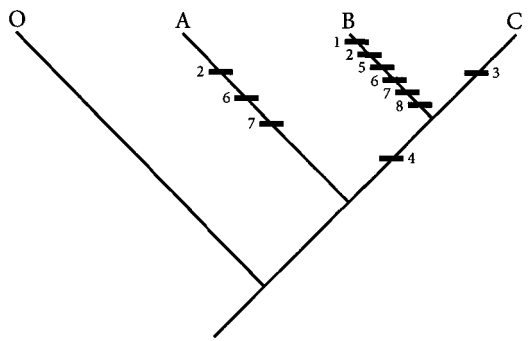


FIGURE 7.5 Tree 1 with all character state changes mapped.

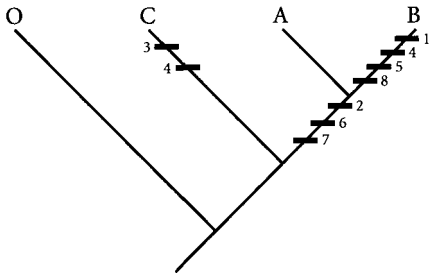


FIGURE 7.6 Tree 3 with all character state changes mapped.

TABLE 7.6 The length of each character on each of the possible trees.

	1	2	3	4	5	6	7	8		
O	0	0	1	0	1	1	0	0		
A	0	1	1	0	1	0	1	0		
B	1	1	1	1	0	0	1	1		
C	0	0	0	1	1	1	0	0	**Total length**	
Length on tree 1	1	2	1	1	1	2	2	1	11	
Length on tree 2	1	2	1	2	1	2	2	1	12	
Length on tree 3	1	1	1	2	1	1	1	1	9	←**Most parsimonious**

We can now apply the same procedures to the other trees. For these same data, tree 2 has a length of 12, whereas tree 3 has a length of 9 (Figure 7.6). This tells us that tree 3 is the most parsimonious tree and is the one that would be preferred under the maximum parsimony optimality criterion.

The length of each tree is a summation of the number of steps required to explain each character on that tree. As Table 7.6 shows, the tree length corresponds to the sum of the length of each of the eight characters.

It is worth highlighting that, although the optimal (most parsimonious) tree has the shortest length overall, it is not optimal for all characters. Character 4 has a length of two on the optimal tree (and on tree 2), but a length of one on tree 1. Character 4 can be said to support tree 1 over trees 2 and 3. However, the totality of the evidence still favors tree 3 over tree 1.

TABLE 7.7 Examples of informative and uninformative characters, with ? used to represent uncertain or missing character states.

Taxon	Parsimony-informative							Parsimony-uninformative						
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	1	0	1	0	?	1	0	1	0	0	0	1	0
3	0	1	1	2	1	0	2	0	1	1	0	0	2	0
4	1	1	1	1	?	1	2	0	1	0	1	1	2	1
5	1	1	2	3	1	1	0	0	1	2	?	2	2	2
6	1	0	2	2	1	?	1	0	1	0	?	3	2	?

A second observation can be drawn from this table. Characters 1, 3, 5, and 8 have the same length on each of the three trees. As a result, they do not contain information that helps choose among trees—they are ***parsimony-uninformative.*** If these characters were deleted, each tree would be four steps shorter than shown, but the rank order and the difference in tree length would be the same. These characters are uniquely derived characters (sometimes called ***autapomorphies***). Because they can be parsimoniously explained as having evolved on a terminal branch of the tree, they do not help tell us which tips share more recent common ancestry. The other four characters are ***parsimony-informative*** in that their length varies among trees. Only parsimony-informative characters have the potential to influence which tree is optimal under the parsimony criterion.

You may wonder whether it is possible to know if a character is uninformative before looking at the length of all the possible trees. Here is a simple rule of thumb: a character is parsimony-informative if there are at least two states that each occur in two or more taxa. Some examples of parsimony-informative and -uninformative characters are shown in Table 7.7.

A JUSTIFICATION OF PARSIMONY

The aim of phylogenetics is to choose the tree that is most likely to be true given all of the observed trait data and our prior understanding of the evolutionary

process. Why would we expect more parsimonious (shorter) trees to be better estimates of the true tree than less parsimonious (longer) trees? Why does a set of data for which tree 1 is shorter than tree B suggest that tree A is a better hypothesis than tree 2?

In Chapter 4 we described a metaphor for the principle of parsimony, in which an emergency call center in a North American city receives two calls in the same day reporting a tiger on the loose. Because reports of tigers are rare, it is logical to assume that the two calls refer to the same tiger. Likewise, if characters change state relatively rarely, then when two tips share the same derived state, it is more likely that the trait evolved once in the tips' common ancestor than that it evolved twice. Thus, taxa sharing a derived character state are *a priori* likely to form a clade. This does not mean that the derived character state absolutely must have evolved once, just that this is the more likely explanation for this character's evolution, taken in isolation.

This principle can be extended to consider all of the characters in a data matrix. Let us start by assuming that the rate of evolution of all characters on the true tree is reasonably low. In that case, the true tree should be able to explain most of the characters' evolution by invoking few evolutionary events, so the length of the tree will be relatively short. An incorrect tree, in contrast, may be able to explain the evolution of some characters without invoking homoplasy, but for most characters we expect them to show some homoplasy on the incorrect tree. Some homoplasy is also expected on the true tree, but the amount of homoplasy should be less than on the incorrect tree. Thus, we expect the true tree to be shorter than incorrect trees.

But what if all traits evolve very rapidly? Should we still expect the true tree to be shorter than the incorrect tree? Because many squirrel sightings occur every day in a typical North American city, the fact that two squirrel sightings occur on the same day is not compelling evidence that the squirrels observed are one and the same. Similarly, if a trait has evolved rapidly, the shared occurrence of a derived character state is not compelling evidence that the taxa with the derived state form a clade. Nonetheless, unless there is a reason to expect different characters to show the same homoplasious pattern, we expect different rapidly evolving characters to support different trees.

If the rate of evolution for all characters in a data matrix is very high, the data should lack a consistent signal favoring one tree over another: that is, all trees should be about the same length. Some simple statistical methods are available to detect cases in which there is no phylogenetic signal in a data set (see Chapter 9).

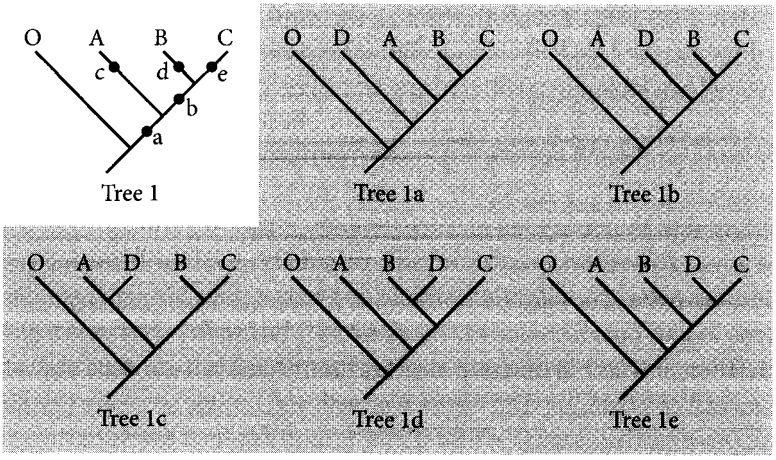
But what if the rate of evolution is low for some characters but high for others? In this case the rapidly evolving characters should tend to yield a noisy pattern that will not strongly favor any one tree over the others. Still, provided that there are enough slowly evolving characters, these will tend to agree with one another and should collectively support trees that resemble the true tree. Thus, parsimony should still tend to point toward the true tree.

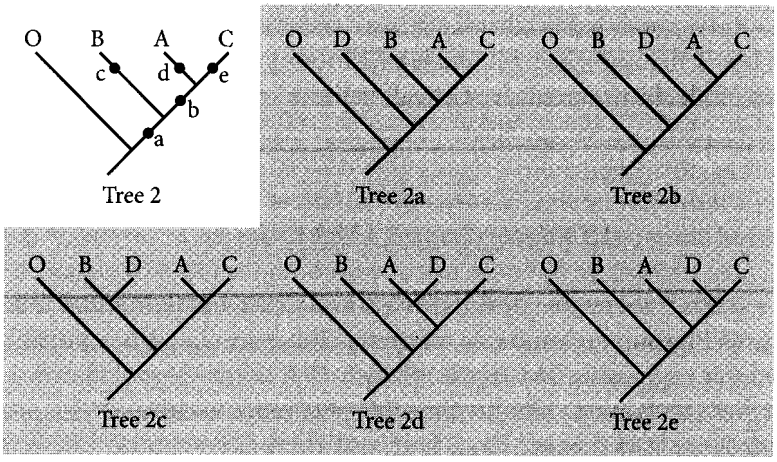
Based on this reasoning, we can see that when the rate of evolution is low, at least for some characters, parsimony is a reasonable tool for inferring phylogenetic trees: shorter trees are more likely to be true than longer trees. Furthermore, tree length provides a crude measure of how much better one tree is than another. Thus, if tree 1 is one step shorter than tree 2 but 15 steps shorter than tree 3, we can say that the data argue against tree 3 more strongly than against tree 2. However, the magnitude of the length difference between two trees is dependent on the particular matrix of characters used. Without other analyses (such as those presented in Chapter 9), we cannot assert that one tree is "significantly" better than another. Although parsimony gives us valuable insights into the trees implied by our data, statistical methods must be applied to determine whether the data convincingly favor some topologies over others.

FINDING OPTIMAL TREES

In the example above, we dealt with a simple case involving just four taxa. With three ingroup taxa and one outgroup, there are only three possible fully resolved tree topologies. This made it easy to determine the parsimony score of each of these trees and to identify the most parsimonious tree.

Suppose instead that we have four, not three, ingroup taxa (five taxa in all). Because the fourth taxon could be added to any of five branches on the three possible trees for three ingroups, there are 15 possible rooted topologies (Figure 7.7). In turn, each of these 15 trees has seven places to add yet another ingroup taxon, meaning that there are $7 \times 15 = 105$ possible rooted trees for five ingroups, and so on. For the mathematically inclined, if we assume one outgroup and n ingroup taxa, the number of rooted tree topologies is $(2n-3) \times (2n-5) \times (2n-7) \times \ldots \times 3 \times 1$. This can also be written: $(2n-3)! \,/\, [2^{n-2} \times (n-2)!]$. Table 7.8 lists the numbers of tree topologies for cases with even more taxa. As you can see, the number of possible trees increases very rapidly as the number of taxa increases. When you get to 52


O A B C
c d e
b
a
Tree 1
O D A B C
Tree 1a
O A D B C
Tree 1b
O A D B C
Tree 1c
O A B D C
Tree 1d
O A B D C
Tree 1e


O B A C
c d e
b
a
Tree 2
O D B A C
Tree 2a
O B D A C
Tree 2b
O B D A C
Tree 2c
O B A D C
Tree 2d
O B A D C
Tree 2e

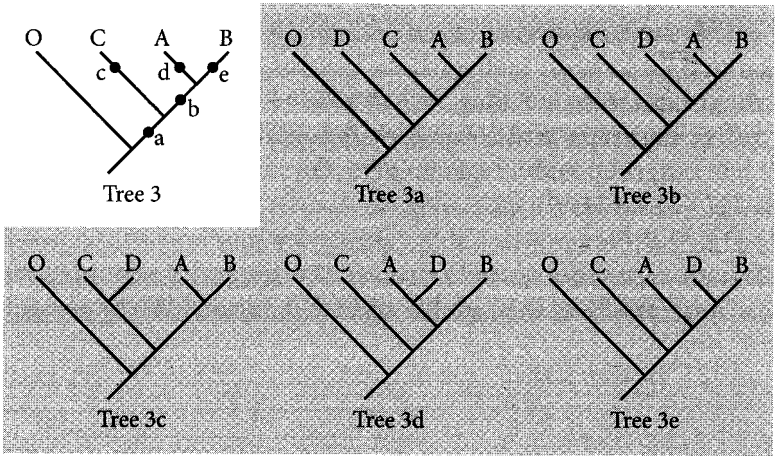

O C A B
c d e
b
a
Tree 3
O D C A B
Tree 3a
O C D A B
Tree 3b
O C D A B
Tree 3c
O C A D B
Tree 3d
O C A D B
Tree 3e

TABLE 7.8 The rapid increase in the number of rooted tree topologies with increasing numbers of taxa

Number of ingroup taxa (with a single outgroup)	Number of possible topologies
2	1
3	3
4	15
5	105
6	945
10	34,459,425
20	~8,200,794,532,637,890,000,000
50	~27,529,213,532,835,600,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000
100	~33,499,038,543,080,200,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000

For example, suppose we have 12 taxa in a data set and have already found one 12-taxon tree of length 712. If we calculate the length of a certain 10-taxon tree and find that it had a length of 713, we would know that any 12-taxon tree that could be pruned to yield this 10-taxon tree must be less parsimonious than the 12-taxon tree that we have already found. Therefore, all trees that are derived from the pruned tree can be excluded without even calculating their length. Depending on the structure in the data, branch-and-bound can usually find the most parsimonious tree while only calculating the length of a subset of trees. However, even branch-and-bound becomes impractical with more than 20 taxa. So what can be done for yet larger numbers of taxa?

Computer scientists have developed ***heuristic search*** algorithms that allow one to analyze indefinitely large data sets. These programs are not guaranteed to find the optimal tree, but they usually do so and even when they do not, the optimal tree overall is expected to be quite similar to the best trees found. To explain how heuristic algorithms work, it will be helpful to introduce the concept of tree space.

As discussed in Chapter 3, one measure of the similarity of two tree topologies is the number of rearrangements needed to convert one tree into the other. Among the several possible rearrangement methods, we introduced one: subtree pruning and regrafting (SPR).

Now imagine a space in which all possible trees are cleverly laid out such that each tree is placed adjacent to all those trees from which it is one SPR rearrangement away. This tree space is multidimensional: each dimension being the distance of one specific tree to all the other trees. However, for the sake of visualization let us pretend that tree space could be flattened into two dimensions.

Now imagine an additional dimension: a measure of tree quality. In a parsimony framework, tree length is a measure of how well a tree explains a set of data. Each point in tree space corresponds to a tree with a definite tree length for the data we are analyzing. Somewhere on this space there must be one or more trees that are shorter than all the rest—the most parsimonious trees. All other trees are a certain number of steps less parsimonious: some are one step longer, some are two steps longer, and so on. The worse the tree, the lower its "altitude" (Figure 7.8). Thus, our objective is to search through tree space to find the highest peak, which corresponds to the set of most parsimonious trees.

Tree space is not infinitely rugged. Two adjacent trees cannot be very different in altitude (length) because a single rearrangement of a tree cannot greatly change the length of all characters. This means that the best (shortest) trees will

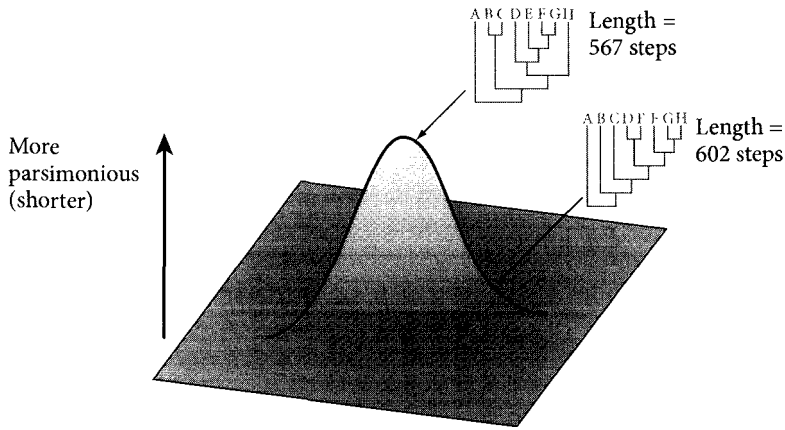


FIGURE 7.8 Visualization of tree space, where the best (shortest) trees sit at the peak.

tend to be adjacent to trees that are almost as good and the worst (longest) trees will generally be adjacent to other bad trees. This fact allows computer programs to search out peaks (short trees) without having to survey all the trees in tree space.

The approach that phylogenetic computer programs usually use when looking for the optimal trees is a hill-climbing algorithm. The program grabs a starting tree (there are clever ways to start searches on decent trees to speed up the analysis) and calculates its length. It then "visits" all the adjacent trees by making all the possible rearrangements to the starting tree, and for each adjacent tree it calculates the length. If the initial tree is shorter than all adjacent trees, it is a peak and the search stops. If it is not a peak, then the computer identifies the shortest of the adjacent trees and then "moves" to that tree (Figure 7.9). It then looks at all of *its* adjacent trees, and so on. By reiterating this procedure, the algorithm is guaranteed to identify a tree or set of related trees that are more parsimonious (= higher) than their surrounding trees. Thus, instead of calculating the number of steps on every possible tree, a heuristic search moves toward the most parsimonious tree or trees by wandering through tree space to successively shorter trees.

Heuristic searches can be adjusted in various ways to make them run faster and to have a higher probability of finding the global optimum in cases of "rugged" tree spaces that have local optima (the shorter peaks in Figure 7.9).

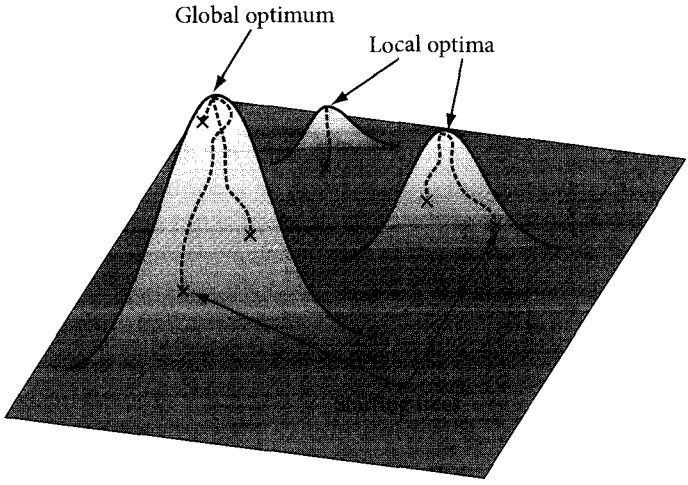


FIGURE 7.9 **Searching a tree space with multiple optima.**

For example, it is common to repeat heuristic searches hundreds or thousands of times initiating with a slightly different starting tree. This procedure would be analogous to finding the highest point on an island by parachuting explorers all over the island with instructions to walk uphill until they reached a peak and to then report the peak's altitude. If most of the explorers met on the same peak, you would be more confident that the island had a conical form and that the true peak had been found. If instead each explorer found a different peak, you should worry that the landscape is a jagged space whose global peak had not yet been found.

Through these and other procedures, computer programs have become able to apply the parsimony criterion to data sets with more than 1000 tips. In consequence, computer power is no longer a major impediment to conducting phylogenetic analysis using parsimony.

PARSIMONY ANALYSIS OF THE CARNIVORAN MORPHOLOGICAL DATA

Having introduced the principles of parsimony and the concept of tree space, we can return to the carnivoran data set to find the most parsimonious trees.

TABLE 7.9 Morphological data for Carnivora

	1 (4)	2 (21)	3 (32)	4 (45)	5 (52)	6 (54)	7 (56)	8 -	9 (59)	10 (60)	11 (61)	12 (62)	13 (40)	14 (50)	15 (51)	16 (1)	17 (2)	18 (3)	19 (24)	20 (26)
Outgroup	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Cat	0	1	0	1	0	0	1	1	1	0	0	0	1	1	1	0	0	0	0	0
Hyena	0	1	0	1	0	0	1	0	1	0	0	0	1	1	1	1	0	0	0	0
Civet	0	1	0	0	0	0	0	0	1	0	0	0	0	0	1	1	0	0	0	0
Dog	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Raccoon	1	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0
Bear	1	0	0	0	1	1	0	0	0	1	0	0	0	0	0	0	1	1	0	1
Otter	1	0	0	0	1	0	0	0	0	1	0	0	1	0	1	1	1	0	0	0
Seal	1	0	1	0	1	1	0	0	0	1	1	1	0	1	1	1	?	0	1	1
Walrus	1	0	1	0	1	1	0	0	0	1	1	1	0	0	1	1	0	1	1	1
Sea lion	1	0	1	0	1	1	0	0	0	1	0	0	0	1	1	1	0	1	1	1

The data matrix shown in Table 7.9 includes the 13 characters listed earlier plus 7 more characters from Wyss and Flynn (1993). In case you want to refer back to the original study, the numbers in parentheses correspond to the character numbers used in that study. Character 8, the presence or absence of retractile claws, was added to provide an example of a parsimony-uninformative character.

Using the branch-and-bound algorithm implemented in the computer program PAUP* (Swofford 2002), we find that there are two equally most parsimonious trees for these data, requiring 30 character state changes to explain the 20 morphological characters. The trees' consistency index (equal to the average CI of the 20 characters) is 20/30, or 0.67. As shown in Figure 7.10, these trees differ only in the resolution within the pinnipeds. The information common to both trees can be shown in a strict consensus tree (Chapter 3), a tree that contains only those clades present in all equally most parsimonious trees, as shown in Figure 7.11. The strict consensus tree has a polytomy where the two equally most parsimonious trees disagree.

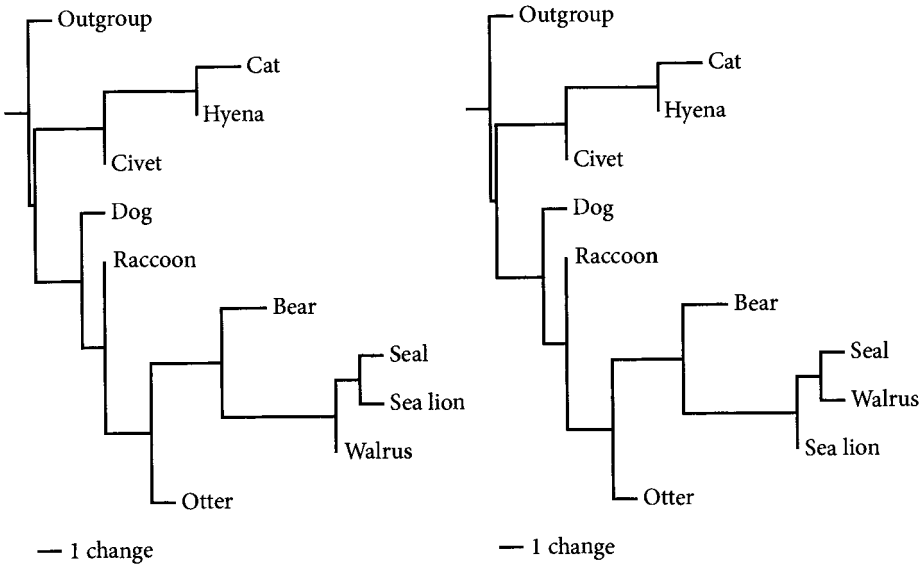


FIGURE 7.10 The two most parsimonious trees for the carnivoran morphology data set. Trees were rooted with the outgroup. These trees differ only in the resolution within the pinniped clade (seal, walrus, sea lion). To indicate the correct rooting, a short internal branch has been added at the base of the tree.

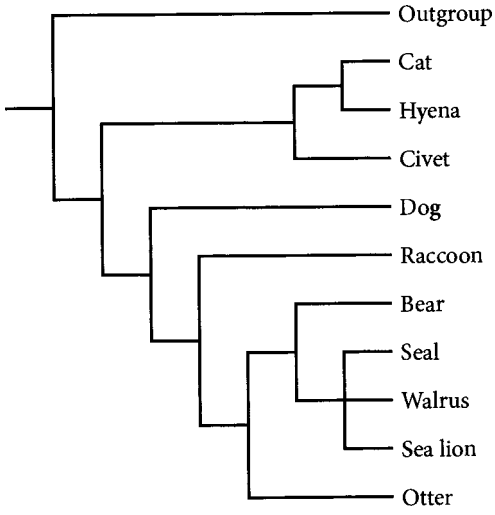


FIGURE 7.11 Strict consensus of the two most parsimonious trees for the carnivoran morphology data set. This shows that the only point of uncertainty is within the pinniped clade (seal, walrus, sea lion).

PARSIMONY ANALYSIS OF DNA SEQUENCE DATA

Once molecular sequence data became available in the 1980s, scientists began using these data for phylogenetic inference using maximum parsimony. As outlined in Chapter 8, most modern phylogenetic studies of molecular data utilize maximum likelihood or Bayesian inference methods in addition to or instead of maximum parsimony. Nonetheless, we believe it will be useful to describe parsimony analysis of DNA sequence data because (a) this accurately summarizes the historical development of the field of phylogenetics and (b) doing so will allow us to introduce some themes and concepts that will come into play when we introduce maximum likelihood approaches.

The raw data used for phylogenetic analysis are DNA sequences obtained for each taxon. Two such sequences are shown below. You will see that they are slightly different in length, having different numbers of bases. In order to proceed to use them for phylogenetic analysis, the first step is ***sequence alignment***: aligning sequences to one another so that a nucleotide position in one is matched with the homologous position in other sequences.

Taxon 1:

```
CGTTTATGGTGACGGAGCCGGGGGAGGTAGCACGTGGCAAAAAGAACGGCCTCGATTATCTCTTCCAT
CTTTACGAACAGTGCCGGGAGTTCTTGATTCAAGTCCAAAACATCGCCAAGGAGCGCGGCGAAAAATG
CCCCACCAAGGTAACAATAGAAACAAATCTATTTTTAATGTTTCTTAAGTAAAATTTTGAATTCAAGC
TCCGTAAATGAATGAAAATATGAGAAATATCCTGTTTTTGATCCGATTCTCATGGAAAAATATGAAAC
TAGGATAGTTTTTGCATGGTGCACGAGGTTTGACACGGGACTAGCTGTAAAAACAAGGCTGTCTCTGT
TAGAATCTTAGAACTGGACCAGCCCTCCCATTAAAGCTAGGGTTTCTAGCCCATGAAAATGTGACAAC
TCAGGTACGGGGAGGAATGGAGTCTGAAAACTTGGGACATGTATGTCTAAATTTTTGCAGAGTAAGGT
CCCCTCCGCCCCAAAAGGTTGTACTTTTGTCTTTAAAGACTTTACTGTCTTCCTTTCTGAAGCCTCGT
TTTCCCTGTCCGGTTTAGCTGAGGTGGCGTGACCCTAATACGACAGCTCCACCAYTTTTGGATCCTAA
TCTTATTGCTTATACAGGTGACCAACCAAGTTTTCAGATATGCTAAGAAGGCTGGGGCGAGCTACATT
AACAARCCCAAAATGMGCCATTACGTCGGCAGGA
```

Taxon 2:

```
TCACCCACGACCGTTCATGGTGACGGAGCCGGGGGAGGTAGCAAGTGGCAAAAAGAACGGCCTCGATT
ATCTCTTCCATCTTTACGAGCAGTGCAGGGAGTTCTTGATTCAAGTCCAAAACATCGCCAAGGAACGC
GGCGAAAAATGCCCCACGAAGGTAACAATAGAAACAAATCTATTTTTAATGATTCTTAAGTAAAATTT
TGAATTCAAGCGTAAATGAATGAAATATGAGAAATATCCTGTTTTTGATCCGATTCTCATGGAAAAAT
ATGAAACTAGGATAGTTTTTGCATGGTGCACGAGGTTTGACACGTGACTAGCTGTAAAAACAAGGCTG
TCTCTGTTAGAATCTTAGAACTGGACCAACCCTCCCATTAAAGCTAGGGTTTCTAGCCCATGAAAATG
TGACAACTCAGGTACGGGGAGGAATGGAGTCTGAAAACTTGGGACATGTATGTCTAAATTTTTGCAGA
GTAAGGTCCCCTCCGCCCCAAAAGGTTGTACTTTTTGTCTTTAAACACTTTACTGTCCTCCTTTCTGA
AGCCTCGTTTTCCCCTGTCCGGTTGAGCTGAGGTGGCGTGACCCTAATACGACAGCTCCATTGGATCC
TAACCTTGTTACTTATACAGGTGACCAACCAAGTCCTCAGATATGCTAAGAAGGCTGGGGCGAGCTAC
ATTAACAAACCCAAAATGCGCCACTATGTC
```

Recall that a data matrix is composed of characters that are shared by taxa but potentially differ in state. For example, the character hair may adopt such character states as white, brown, or black. For DNA sequences, the character is the nucleotide position (numbered 1, 2, 3, etc.) and the character states are the nucleotides (A, C, G, and T). It is critical, therefore, that nucleotides in each taxon be assigned to the correct positions. Sequence alignment involves sliding the sequences over one another and inserting gaps, guided by the sequences themselves. Sequence alignment, done properly, can pose major computational challenges and has become a very technical subject. Here, we will just summarize the underlying issues and point to some additional resources.

A DNA molecule is a physical structure with nucleotides in a specific linear order. In the simple case in which the only kind of mutations are base substitutions, each nucleotide position in one taxon would be homologous to a nucleotide position at the same place in the sequence of another taxon: position 1 in taxon A will be homologous to position 1 in taxon B, position 2 to position 2, and so on. If we write out the two sequences, the homologous positions are aligned above one another. In the example below, the sequences from two closely related taxa are the same length but have some differences (shaded) due to base substitutions.

```
Taxon A:  G T A T T G A C C A C T G A C T A G C A T
          | | | | | | | | | | | | | | | | | | |
Taxon B:  G C A T T A A C C A T T G T C T A G C A A
```

If the only kind of mutations were base substitutions, having found the homologous genes you would merely need to line up one homologous position and the rest of the alignment would be trivial. However, sequences are subject to additional kinds of mutation: deletions, insertions, inversions, and translocations. Of these, insertions and deletions appear to be the most common.

A ***deletion*** involves the removal of one or multiple continuous bases. Deletions may be due to errors during DNA replication, but can also happen due to imperfect DNA repair following damage, unequal crossing-over during recombination, or the action of mobile genetic elements. Deletions can be as short as one base pair or as long as thousands. When deletions happen, nucleotide positions in one sequence may lack any homolog in another sequence. The missing positions can be marked with a dash. For example, here is a case where a sequence experienced a five base-pair deletion relative to its ancestor.

```
Ancestor     G T A T T G A C C A C T G A C T A G C A T
             | | | | | | | | | | | | | | | | | | |
Descendant   G C A T T - - - - - T T G T C T A G C A A
```

The same mechanisms that cause deletions (errors during replication and recombination, DNA damage, and mobile genetic elements) can cause the ***insertion*** of DNA sequences into a strand. When insertions happen, new bases emerge that have no homologs in the ancestral sequence (they may be copied from somewhere else in the genome, but we rarely know this). The lack of homologs in one taxon can be indicated, again, with a dash. In the following example, a sequence has experienced a three base-pair insertion relative to the ancestral sequence.

```
Ancestor     G T A T T G A C C - - - A C T G A C T A G C A T
             | | | | | | | | | | | | | | | | | | | | | |
Descendant   G C A T T A A C C A C C A T T G T C T A G C A A
```

In practice we generally do not have access to ancestral sequences. When we find a gap in one sequence relative to another sequence, we do not know whether there was an insertion or deletion. In light of this ambiguity, the processes that generate gaps are often called insertion/deletion events, or indels (see also Chapter 4).

The process of sequence alignment aims to align homologous positions based on the true history of sequence evolution. Alignment is, thus, properly viewed as a problem of historical inference. Furthermore, because base substitutions and indels occurred along the branches of the gene tree, sequence alignment and tree inference are really two aspects of the same problem. Therefore, in the ideal world, we would have computer programs that could take raw, unaligned sequences and search for trees that could simultaneously account for the bases in the sequences and their indels. A few programs do conduct such combined alignment and phylogenetic inference. However, the problem is so computationally challenging that most phylogenetic analyses separate the two problems: first generating an alignment and then provisionally accepting that alignment as the basis for phylogenetic inference. Combined alignment and tree inference is, however, likely to become more common over the next decade.

To get a feel for how sequence alignment can be conducted free of a phylogeny, see if you can align the following pair of sequences.

```
A T G A C C T G G C G G C T T T A
A T G T G G A T A T G G C A T T A
```

You might conclude that these sequences are already well aligned, that there were seven substitutions affecting the shaded positions.

While this might be the best alignment, it is worth considering alternatives that can also explain these data through the addition of indel events. For

example, you could align these same two sequences by invoking five indels and no substitutions, or two substitutions and two indels, as shown below. Remember that an indel event can involve any number of bases, so whether it is one dash or four, it still counts as one indel.

Five indels:

```
A T G A C C T G G - - - - C G G C T - T T A
A T G - - - T G G A T A T - G G C - A T T A
```

Two substitutions and two indels:

```
A T G A C C T G G - - - C G G C T T T A
A T G - - - T G G A T A T G G C A T T A
```

To choose between these alignments, we need to ask ourselves whether it is more likely that there were five indels, or two substitutions and two indels. Data from molecular biology would say that base substitutions are generally more frequent than indels (especially in coding genes). Thus, we would tend to reject the first alignment because it invokes more indels than substitutions in addition to invoking more total events (five versus four).

This allows us to state a rule that is applied in almost all sequence alignment programs: invoke indels only when the number of base substitutions avoided is greater than the number of indel events implied. Indeed it is normal to set the ***gap penalty***, the threshold for the number of base substitutions avoided before an indel is inferred, higher still: gap penalties of 3 to 20 are common. Additionally, most computer programs impose an extra cost for longer gaps or gaps at the ends to avoid useless alignments such as the following, which invokes no base substitutions at the "cost" of two indels:

```
- - - - - - - - - - - - - - - - - - A T G A C C A G T A C G G C T T T A
A T G A T C G A T A T G G C A T T A - - - - - - - - - - - - - - - - - -
```

It is probably clear that alignment is easiest and most certain when both base substitutions and indels are rare. This is because conserved parts of the sequence provide a framework for identifying the position and size of indel events. For example, below are two true alignments. Which do you think we would be more likely to correctly infer?

```
A T G A - - - T G C A G C T T T A G G T A
A C A A C A G T A C G A - - C T A C - C A

A T G A C C A G T A C A G - T T T A G T T
A C G A C C - - T A C G G C T T C A G T A
```

The answer is the second one. While the number of indels is similar, the many extra base substitutions in the top case would make it very hard to identify the true alignment with confidence.

Pairwise alignment considers just two sequences at a time, whereas multiple alignment includes sequences from many taxa to obtain an entire aligned data matrix. A pairwise alignment is relatively simple for a computer to determine, even when a complex set of penalties is implemented. Multiple alignments are, however, disproportionately more difficult. As the number of sequences being aligned increases, the number of possible alignments goes up exponentially. Multiple alignment algorithms should allow the placement of gaps in one sequence to influence the placement of gaps in other sequences. This is because, when gaps in two species are in the same position, they can be attributed to a single indel occurring somewhere on the gene tree. However, detecting shared indels is not always easy for a computer program. As a result, although multiple alignment programs (e.g., CLUSTAL, MAAFT, MUSCLE, TCOFFEE, FAS) provide a good starting point, they usually need to be examined and adjusted by eye. To illustrate this, Figure 7.12 shows a problematic portion of an alignment that was returned by CLUSTAL and a manually edited version of the same. You

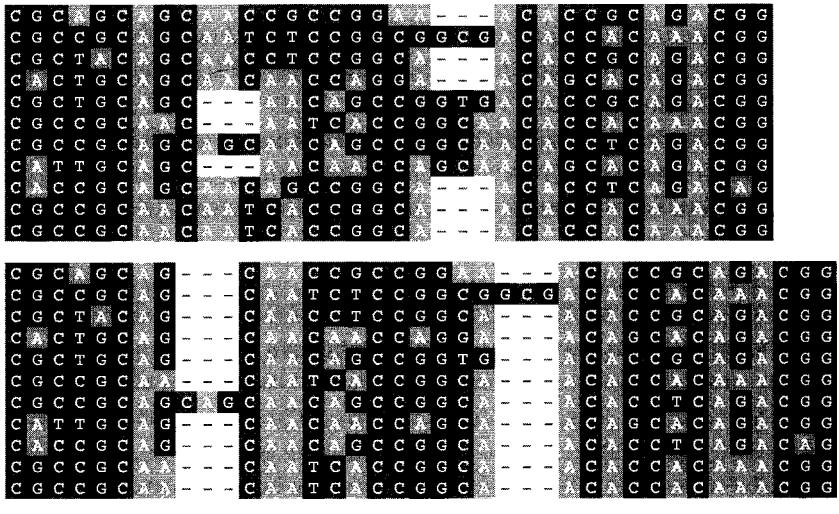


FIGURE 7.12 **Comparison of a computer and manual ("eyeball") alignment of the same data set.** The upper alignment was generated by a commonly used multiple alignment program, whereas the lower alignment was generated by hand.

might notice that not only does the second alignment imply a simpler mutational history, but also the human editor could take account of the codon structure of this gene (something that few computer programs keep track of) so as to keep both gaps in the same reading frame.

Given that sequence alignments will vary depending on how they were generated (which algorithm, what penalties, and whether they were manually edited), you might worry that phylogenetic analysis of DNA sequences is invalid. Actually, the problems are less than they may seem. Even if some part of a sequence is hard to align unambiguously, many regions can often be aligned confidently. The regions that are aligned correctly will tend to be composed of characters that provide consistent support for the same tree, whereas characters in regions that have been misaligned will tend to conflict with one another—they will constitute phylogenetic noise, similar to very rapidly evolving traits. This means that reasonable phylogenetic conclusions can often be obtained even when the alignment is imperfect. Nonetheless, it is wise to obtain a sense of how one's conclusions depend on the alignment. This is typically done by reconstructing the phylogeny using several different alignment schemes to see whether the major phylogenetic conclusions change. Finally, it is worth considering the use of one of several computer programs that treat sequence alignment and tree estimation as a single problem.

Having aligned DNA sequences, it is straightforward to analyze them using maximum parsimony (or with the methods described in Chapter 8). In its most basic form, parsimony analysis of DNA sequences counts all character state changes the same, regardless of which character is involved and what kinds of substitutions are invoked. Gaps are usually treated as missing data, but sometimes the inferred indel events are treated as additional characters.

To provide a concrete example, let us consider some DNA sequences obtained for representatives of the Carnivora. Sequences are available for a 1116 base-pair region of the transthyretin 2 gene (Flynn and Nedbal 1998) for the living species (Table 7.4). These sequences were aligned by eye and were then entered into a computer program, PAUP* (Swofford 2002), which searched for the most parsimonious tree. A branch-and-bound search yielded two trees of length 790 steps, which differed only in the placement of hyena (sister to civet or to cat). The strict consensus of the two is shown in Figure 7.13.

If you compare Figures 7.13 and 7.11, you will see that the trees obtained from morphological and molecular data are similar. Both data sets support a monophyletic pinniped group (seals, walruses, sea lions) and a division of the other living carnivores into two major subclades: feliforms (cats, hyenas, civets)

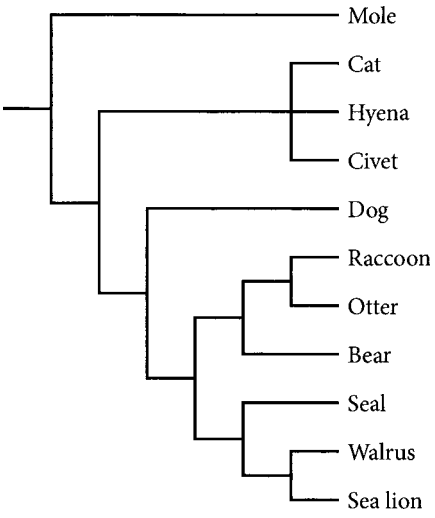


FIGURE 7.13 Consensus of the two most parsimonious trees for the carnivoran molecular data. Trees were rooted with the outgroup, mole.

and caniforms (dogs, raccoons, bears, otters, and pinnipeds), with dogs being the sister taxon to a clade comprising the remaining caniforms. This agreement is significant. If these data were not the result of evolution along a tree, the probability that the two data sets would yield trees this similar is less than 1 in 2000 (0.000375). The fact that both kinds of data yield such similar trees provides concrete evidence in support of the claim that these data are the result of descent along the same evolutionary tree. As discussed in Chapter 2, the agreement among independent phylogenetic data sets supports the hypothesis of common ancestry.

Although the trees inferred from morphological and molecular data are remarkably similar, they also have some differences. One difference is the sister group of the pinnipeds. The morphological data suggest that the bear lineage is the sister group (Figure 7.10), whereas the DNA sequences suggest that a clade composed of bear, otter, and raccoon is in this position. Such differences are best understood as being due to imperfect phylogenetic inference. One or the other or both trees are presumably incorrect in some details. While some conclusions, such as the fact that pinnipeds are embedded within the caniforms, are well demonstrated by these analyses, other results would have to remain uncertain pending the collection of more data. In fact, abundant

additional data have been collected on carnivoran phylogeny. If you want to know more, you can gain access to the literature by consulting Agnarsson et al. (2010).

CHARACTER AND CHARACTER STATE WEIGHTING

When we do phylogenetic analysis using parsimony (or another method), each character provides an independent piece of evidence of the actual evolutionary past. Drawing an analogy to forensics, another historical science, each character can be equated with a different piece of evidence collected at a murder scene. For example, character 1 might be the position of the body, character 2 might be the location of a bullet hole in the wall, and character 3 might be a scrap of paper in the victim's hand.

Until now, we have made the assumption that all changes are counted equally when deciding which trees are most parsimonious. This approach is called ***equally weighted parsimony*** or ***Fitch parsimony***, because it resembles a model proposed by Walter Fitch (Fitch 1971). Fitch parsimony is analogous to a forensic investigation in which all pieces of information are assigned equal weight in testing the innocence of a defendant. However, while many pieces of forensic evidence might influence one's belief in the guilt or innocence of a defendant, some pieces of evidence could be more compelling than others. For example, we are more likely to return a guilty verdict if the defendant's fingerprints were on the murder weapon than if a car matching the defendant's was seen near the site of the crime.

Applying this reasoning to phylogenetic inference, some characters ought to provide more reliable information about phylogeny. If some characters or kinds of character state change are less likely to show homoplasy than others, they are more likely to be consistent with the true tree and should provide more reliable phylogenetic information. What we need to use is a more flexible version of parsimony, called ***generalized (or weighted) parsimony***, in which we give more weight to those characters that we expect to provide more reliable information on phylogeny. Generalized parsimony allows more detailed prior knowledge of trait evolution to yield a more accurate assessment of the phylogeny. It is worth exploring the basics of generalized parsimony as a way to become more familiar with the logic of phylogenetic analysis, and as a useful lead-in to phylogenetic analysis by maximum likelihood.

The main reason why characters might differ in their tendency to show homoplasy is because they evolve at different rates. Because parsimony is most effective when rates of evolution are low (see A Justification of Parsimony earlier in this chapter), traits evolving more slowly provide more reliable phylogenetic evidence. Different kinds of characters are often expected to evolve at different rates. For example, gene sequences often include both slowly evolving regions (e.g., conserved domains, coding regions) and more rapidly evolving regions (e.g., introns, third codon positions). To assign all characters equal weight even when some provide more reliable evidence than others could give weak characters more weight than they deserve.

Returning to the small, hypothetical data set shown in Table 7.5, suppose that we judged character 4 to be five times as informative as the other characters in the matrix. To reflect this, we could count any change of character 4 as equivalent to five changes of the other characters. In generalized parsimony, the score of a tree is no longer simply the number of changes needed to explain the data but a sum of the cost of each character's evolution, where cost is the product of the character's length (number of steps) and its weight. Table 7.10 shows

TABLE 7.10 Tree scores when character 4 is assigned a weight of 5

	1	2	3	4	5	6	7	8			
O	0	0	1	0	1	1	0	0			
A	0	1	1	0	1	0	1	0			
B	1	1	1	1	0	0	1	1			
C	0	0	0	1	1	1	0	0			
Weight	1	1	1	5	1	1	1	1	**Total length**	**Total cost**	
Cost of tree 1	1	2	1	5	1	2	2	1	11	15	**←Most parsimonious**
Cost of tree 2	1	2	1	10	1	2	2	1	12	20	
Cost of tree 3	1	1	1	10	1	1	1	1	9	17	

the length and cost for each of the three possible trees (see Figure 7.3). The scores of all trees have gone up relative to Fitch (equally weighted) parsimony. However, whereas the scores of trees 2 and 3 have increased by eight (because two changes of character 4 are needed) the score of tree 1 has increased only by four. As a result, tree 1 is now the most parsimonious.

To provide a concrete example of generalized parsimony, let us consider two weighting schemes that we might consider applying to the carnivoran morphological data. First, imagine that you believed that gaining or losing external ears (pinnas) during evolution is rare and decided to reflect this by assigning a weight of 2 to character 11, while all other characters had a weight of 1. Rerunning the parsimony analysis results in finding just one most parsimonious tree with a length of 31. This tree resembles one of the two optimal trees from the flat-weighted analysis (Figure 7.10), the one in which the seal and walrus, which both lack external ears, form a clade.

Now, let us consider an alternative weighting scheme where we double the weight of all tooth characters (characters 1, 3, 4, 13, 14, 15, and 18). In this case there is again a single most parsimonious tree with a length of 42. This tree corresponds to the other tree that was found in the flat-weighted analysis (Figure 7.10), the one in which seals and sea lions form a clade. This illustrates that changing the weight of characters can change the conclusions, although in this case the impact is relatively minor.

Generalized parsimony analysis is flexible enough to accommodate another kind of differential weighting, called ***character state weighting***. Instead of assigning an elevated or lowered cost to all evolutionary changes occurring in a character, one applies different weights to particular character state transitions. Character state weighting is best represented with a step matrix, which shows the cost of transitions between each possible pair of states. For example, Table 7.11 shows a step matrix that corresponds to equally weighted (or Fitch) parsimony for DNA sequences. As you can see, all state changes are assigned the same cost.

With DNA sequence data, bases A and G are chemicals called purines, and C and T are pyrimidines. This matters because mutations within base-types, called ***transitions***, happen more frequently than do mutations between base-types, called ***transversions***. Because transversions occur less frequently, they should be less prone to homoplasy and should be assigned a higher weight. Table 7.12 shows a step matrix that assigns twice the cost to transversions as to transitions. This means that homoplastic transversions exert a greater cost for

TABLE 7.11 Step matrix corresponding to Fitch parsimony with all character state changes receiving the same weight

		To:			
		A	C	G	T
From:	A	0	1	1	1
	C	1	0	1	1
	G	1	1	0	1
	T	1	1	1	0

TABLE 7.12 Step matrix for 2:1 upweighting of transversions relative to transitions

		To:			
		A	C	G	T
From:	A	0	2	1	2
	C	2	0	2	1
	G	1	2	0	2
	T	2	1	2	0

parsimony than do homoplastic transitions. Application of this step matrix to the carnivoran molecular data results in a single most parsimonious tree, which is identical to one of the two from the equally weighted analysis (Figure 7.10). Thus, the general conclusions in this case do not appear to be highly sensitive to changing parsimony costs.

PROBLEMS OF PARSIMONY

As one of the first and most widely used methods for phylogenetic inference, maximum parsimony has led to many phylogenetic discoveries, such as identification of the major branches in the history of flowering plants and resolution of the relationships among humans and other primates. The method has also been tested in the laboratory by generating cultures of viruses with a known phylogenetic history (due to multiplying and periodically subdividing cultures) and using parsimony to infer that history from the viral DNA sequences (Hillis et al. 1992). Nonethless, while we know that parsimony works, it possesses several disadvantages that have led to the increasing use of alternative methods.

First, because parsimony does not take account of branch lengths, it can be led astray when the rate of evolution is high and the true tree has branches that have different lengths (Felsenstein 1978). If the true tree has some very long branches (e.g., due to rapid molecular evolution) and some very short branches

(e.g., due to slower molecular evolution), parsimony will tend to find an incorrect tree. Specifically, parsimony will typically yield a tree that clusters the long branches together. This problem with parsimony is usually called ***long-branch attraction***. To make matters worse, parsimony will tend to support the wrong tree more and more strongly as additional data are collected.

For example, if the tree in the upper panel of Figure 7.14 were true, the resulting data analyzed with parsimony would tend to yield a tree like that shown in the lower panel. Long-branch attraction arises because parsimony

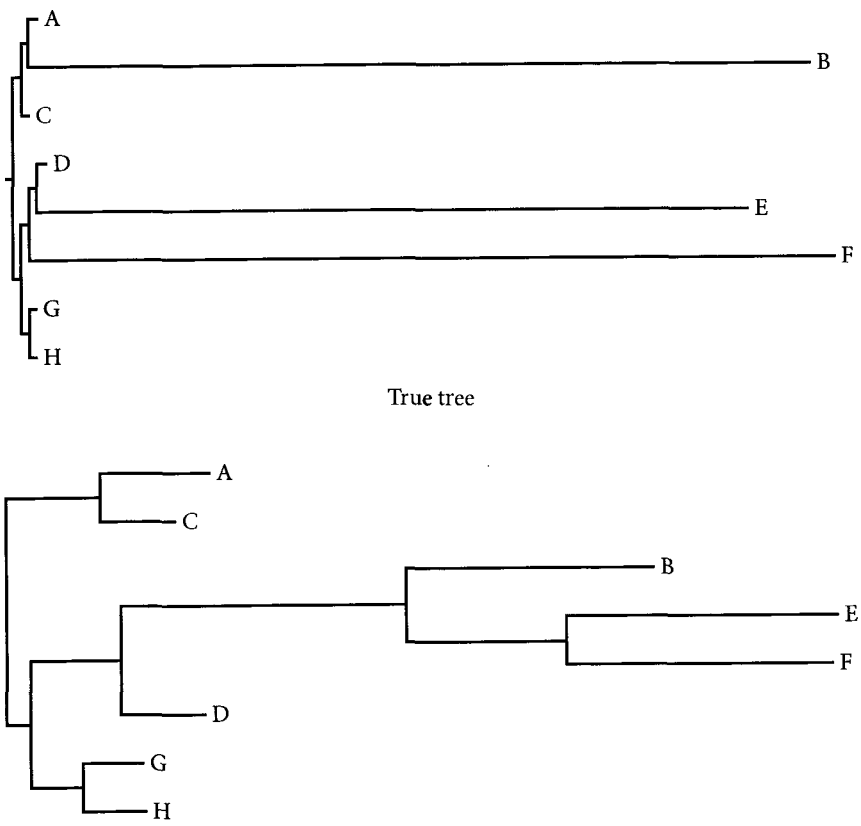


FIGURE 7.14 Long-branch attraction. If the top tree is true, then a typical data set evolving along that tree would, if subjected to parsimony analysis, yield the lower tree.

does not allow information to flow from one character to another about the rate of evolution on different branches of the tree. If a large number of traits are inferred to change on a particular branch, then we ought to allow an increased probability of mapping additional homoplasy to those long branches. However, parsimony assigns changes based only on tree topology, making it unable to detect unusually long or short branches.

Second, parsimony requires us to select a weighting scheme (even if that scheme is equal weighting), and different weighting schemes often affect our conclusions. The problem is that there is no formal method for identifying the most appropriate weighting scheme for a given data set. The best we can do is to examine a range of plausible weighting schemes to assess how robust our conclusions are to the scheme selected. However, this approach reduces our ability to extract all the information in the data, sometimes resulting in poorly resolved estimates of phylogenetic history.

Most scientific journals accept analyses based on parsimony for morphological data, partly because formal mathematical models of morphological evolution are still poorly developed. However, the scientific community generally expects researchers to use maximum likelihood or Bayesian methods (Chapter 8) when analyzing molecular sequence data. Nonetheless, because parsimony is effective for many data sets, is less computationally demanding, and is easier to understand, it is still widely used in educational contexts and for preliminary data exploration.

FURTHER READING

Hennigian inference: Hennig 1966; Felsenstein 1978; Wiley 1981; Brooks et al. 1994
Parsimony/generalized parsimony: Wiley et al. 1991; Swofford et al. 1996
Justification of parsimony: Felsenstein 1981b; Farris 1983, 2000
Tree searching: Maddison 1991; Swofford et al. 1996; Nixon 1999; Quicke et al. 2001
Sequence alignment: Wheeler 1996, 2001; Liu et al. 2009; Morrison 2009
Long-branch attraction: Felsenstein 1978, 1983; Siddall and Whiting 1999; Sanderson and Kim 2000

CHAPTER 7 QUIZ

Questions 1–4. Refer to this data matrix. Assume that taxon H is the outgroup.

	1	2	3	4	5	6	7
A	0	0	0	0	0	0	0
B	0	1	1	0	1	1	0
C	1	1	0	1	1	0	1
D	1	1	0	1	1	0	1
E	1	1	0	0	0	0	0
F	0	1	1	0	0	1	0
G	0	1	1	1	1	1	0
H	0	1	1	1	0	1	0

1. For how many of the scored characters do taxa A and E share the same state?
   a. 0 b. 2 c. 4 d. 5 e. 7
2. For how many of the scored characters do taxa C and E share the same state?
   a. 0 b. 2 c. 4 d. 5 e. 7
3. Applying the principles of Hennigian inference to characters in isolation (i.e., ignoring the rest of the matrix), which of the following characters would suggest that C is more closely related to E than to A?
   a. 1 b. 3 c. 4 d. 5 e. 6
4. Which of the following characters directly contradicts the claim that C and E are more closely related to each other than either is to A?
   a. 1 b. 3 c. 4 d. 5 e. 6

Questions 5–7. Here is a small molecular data matrix. Assume that taxon A is the outgroup.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
A	A	T	G	C	G	G	G	T	C	-	-	-	A	T	A	G	A	T	C	C	A
B	A	T	G	G	G	A	A	C	T	A	G	A	A	T	A	G	A	G	C	C	A
C	A	T	G	A	G	A	A	G	T	A	G	A	A	G	A	G	A	T	C	C	A
D	A	T	G	T	C	A	A	G	A	-	-	-	A	T	A	G	A	G	C	C	A
E	A	T	G	T	C	T	A	T	A	A	G	A	A	T	A	G	A	G	C	C	T
F	A	T	G	T	C	T	T	T	A	A	G	T	A	G	A	G	A	T	C	C	T
G	A	T	G	T	C	A	A	G	G	-	-	-	A	T	G	G	A	A	C	C	A
H	A	T	G	A	C	A	A	G	G	-	-	-	A	G	A	G	A	T	C	C	T

5. What do the dashes in columns 10–12 most likely represent?
   a. Positions occupied by a nonconventional base (neither A, C, G, nor T)
   b. Bases deleted during sequence evolution
   c. Positions where bases were inserted in other sequences
   d. Parsimony-uninformative character states
   e. Sequencing errors

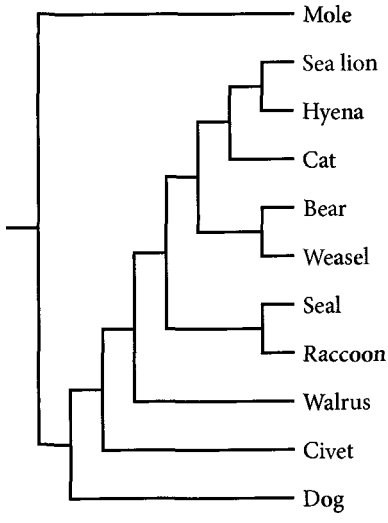
6. Which of the following positions is parsimony-informative?
   a. 3 b. 5 c. 13 d. 15 e. Two of the other answers are correct

7. How would a parsimony analysis be affected by removing the first three positions from the data matrix?
   a. The optimal tree topology will definitely not change; the optimal tree will be three steps longer.
   b. The optimal tree topology will definitely not change; the optimal tree will be three steps shorter.
   c. The optimal tree topology will definitely not change, and neither will its length.
   d. The optimal tree topology might or might not change; the optimal tree length could be longer or shorter.
   e. The optimal tree topology will certainly change; the optimal tree will be shorter.

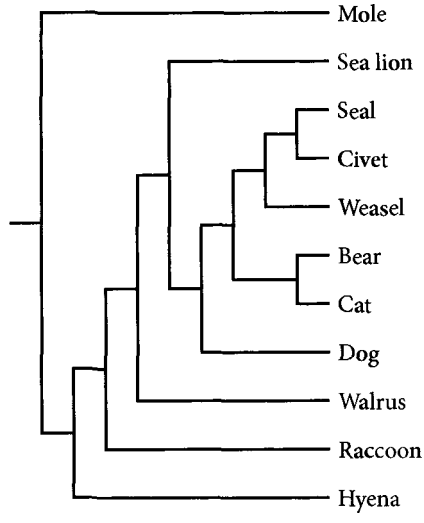
Questions 8–10. Below is a subset of the molecular data that Nedbal and Flynn collected for a phylogenetic study of Carnivora. The sequences come from the transthyretin Intron I. The mole is included as the outgroup.

Mole	G	T	C	A	A	C	A	G	T	C	G	A	C	A	T	T	C	A	T	G	C	T	T	T	T	C	T	T	T	T	G	T	T	T	G	C
Sea lion	A	T	T	C	C	C	G	C	T	C	C	C	T	G	T	T	T	G	T	C	T	A	G	G	C	G	A	T	T	T	A	G	A	G	C	G	C
Walrus	A	T	T	C	C	C	A	C	T	C	C	C	T	G	T	T	T	G	T	C	T	G	G	G	C	G	A	T	T	T	A	G	A	G	C	G	T
Seal	A	T	T	C	C	C	G	C	T	C	C	C	T	G	T	T	T	G	T	C	T	G	G	G	C	A	A	T	T	T	A	G	A	G	C	G	T
Bear	A	C	T	A	A	C	G	C	T	A	T	C	T	G	T	T	T	G	T	C	C	T	G	A	C	G	A	T	T	T	C	C	G	G	C	G	C
Raccoon	A	T	T	G	G	T	G	C	T	A	T	C	T	A	T	G	T	G	C	C	T	T	G	G	C	G	G	T	C	T	C	G	G	G	T	G	C
Weasel	A	T	T	A	A	C	G	C	T	A	G	C	T	A	T	G	T	G	C	C	T	T	G	G	T	A	G	T	C	C	C	G	G	G	C	G	A
Dog	A	T	T	A	A	T	G	G	T	G	T	A	C	C	T	T	T	A	T	C	T	T	T	T	C	G	C	T	T	T	C	C	A	T	C	G	C
Civet	C	C	T	A	A	C	A	G	G	A	T	A	T	A	C	T	G	A	T	G	T	T	T	T	C	A	A	A	T	C	A	G	C	T	C	A	C
Hyena	C	C	T	A	A	C	G	G	G	A	T	A	T	A	C	G	G	A	T	G	T	T	T	T	C	A	A	A	T	C	A	G	C	T	C	A	C
Cat	C	C	C	A	A	C	A	G	G	A	T	A	T	A	C	T	G	A	T	G	T	T	T	T	C	A	A	A	T	T	A	G	G	T	C	G	C

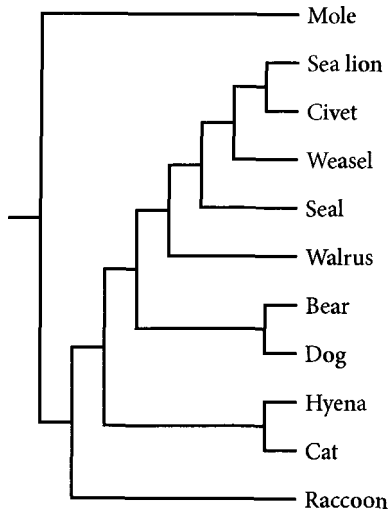
8. Which of the following five random trees is most compatible with the first nucleotide position (shaded), considered in isolation?



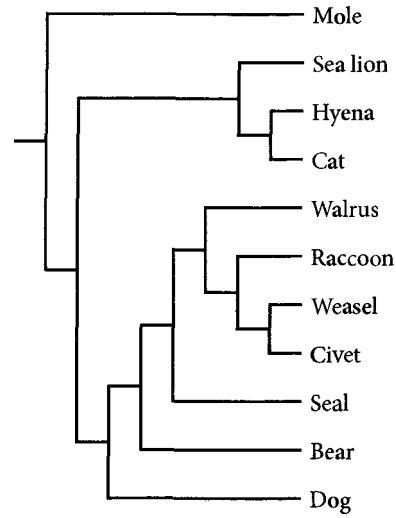
a



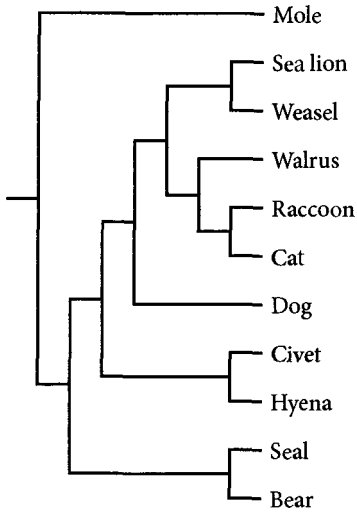
b


Mole
Sea lion
Civet
Weasel
Seal
Walrus
Bear
Dog
Hyena
Cat
Raccoon

c


Mole
Sea lion
Hyena
Cat
Walrus
Raccoon
Weasel
Civet
Seal
Bear
Dog

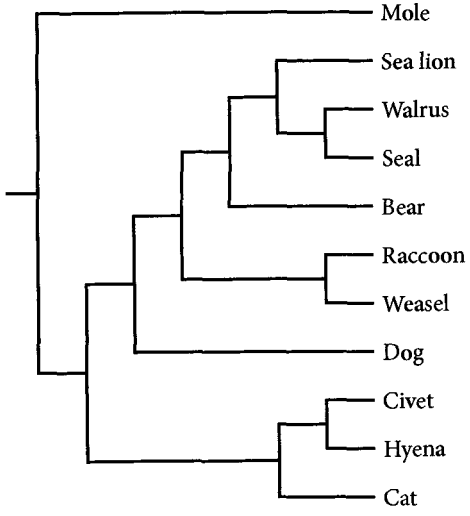
d


Mole
Sea lion
Weasel
Walrus
Raccoon
Cat
Dog
Civet
Hyena
Seal
Bear

e

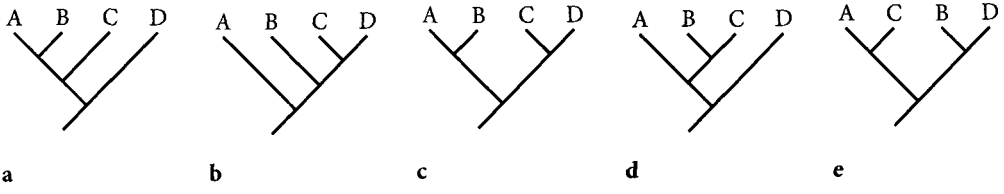
9. The matrix yields a single most parsimonious tree of length 72. What does the number 72 refer to?
   a. It is the number of characters that were analyzed.
   b. It is the number of character states that were analyzed.
   c. It is the number of character state changes needed to explain all the data on this tree.
   d. It is the number of clades in the most parsimonious tree.
   e. It is the number of trees that were considered during the heuristic search procedure before the optimal tree was found.

10. Under the assumption of parsimony, how many changes/steps does one need to explain the first nucleotide position (shaded) in the matrix on this tree?
    a. 1 b. 2 c. 3
    d. 4 e. >4



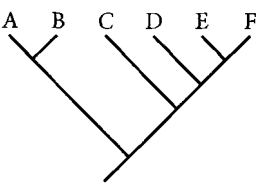
11. Here is a small data matrix. Taxa are labeled A–D in column 1, and the other columns show bases present at five positions in the aligned DNA sequences. Given these data, which of the five trees is most parsimonious?

A	C	A	G	C	G
B	C	T	A	C	A
C	C	A	A	T	G
D	A	T	A	T	A

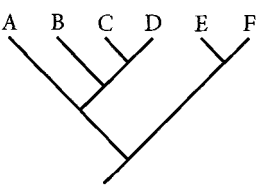


12. Below is a small data matrix and four alternative trees. For each character, calculate the number of steps needed to account for its evolution on each of the four trees. Sum across the characters to determine the overall tree length. Which of the four trees is the most parsimonious?

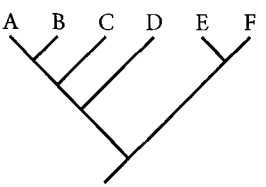
	Character									
Taxon	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	
A	T	G	T	G	A	A	C	A	A	
B	T	G	T	G	A	C	C	A	A	
C	T	G	C	G	G	C	C	T	A	
D	A	G	C	G	G	C	G	T	A	
E	A	A	C	T	A	A	G	T	G	
F	A	A	C	T	A	A	G	C	G	**Total Length**
Steps on tree 1										
Steps on tree 2										
Steps on tree 3										
Steps on tree 4										



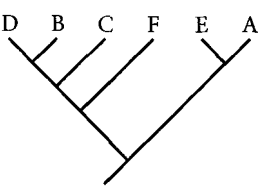
Tree 1



Tree 2



Tree 3



Tree 4

13. Generate (from your imagination) a DNA sequence matrix comprising 10 characters and seven taxa. Design the matrix to have the following features:
    a. All four bases A, C, G, T should be used.
    b. Two characters should include gaps.
    c. Eight characters should have one or two states, two should have three or four states.
    d. Eight of the characters should be parsimony-informative, two should be uninformative.
    e. Six informative characters should be consistent with one another. The other two should conflict with one or more of the other six characters.

Taxon	1	2	3	4	5	6	7	8	9	10
A										
B										
C										
D										
E										
F										
Outgroup										

Draw two rooted trees: one should be the most parsimonious tree for the data matrix and the other should be a less parsimonious tree. For each tree indicate its length (treating indels as missing data).

14. Impose a character or character state weighting scheme to the data matrix in the preceding question and predict how it will change the topology of the most parsimonious tree.

15. Below is a small data matrix. Assuming that taxon A is the outgroup, infer the phylogeny using either Hennigian inference or parsimony (they yield the same result). Based on this tree, is taxon C more closely related to B or D? Based on the data matrix, does C share more traits in common with B or D? How do you explain this discrepancy?

	1	2	3	4	5	6	7
A	0	0	0	0	0	0	0
B	1	0	0	0	0	0	0
C	1	1	1	0	0	0	0
D	1	1	1	1	1	1	1

16. What attributes must characters have or not have in order to be useful for inferring phylogenies using parsimony?

17. The Hennigian method allows one to conclude that some tree topologies are definitely false. Why is this not the case with parsimony?

18. Suppose you are studying a 100-taxon data set. Given the impossibility of calculating the length of every possible tree, can you hope to ever find the most parsimonious tree?

19. It has been argued that equally weighted parsimony is preferable to generalized parsimony because the former does not make assumptions about how characters evolve. What is wrong with this argument?

20. Is it true that parsimony assumes that few if any characters show homoplasy?

CHAPTER EIGHT

Phylogenetic Inference with Distance, Maximum Likelihood, and Bayesian Methods

The previous chapter presented Hennigian inference and parsimony as a means to illustrate the principles underlying phylogenetic analysis in general. However, Hennigian analysis is not in current use, and parsimony, while widely used, is no longer the most common method for phylogenetic inference. In this chapter, we will introduce three other approaches to reconstructing trees. The first, based on evolutionary distances, is frequently used in situations where speed of analysis rather than phylogenetic precision is important. The other two approaches, maximum likelihood and Bayesian analysis, are sophisticated methods that use mathematical models of character evolution to achieve more precise estimates of phylogenetic history. Because of their statistical power, maximum likelihood and Bayesian analysis are the major methods used in phylogenetic research, at least for DNA sequence data. Before describing these three methods, it will be useful to introduce mathematical models of evolution because these provide a basis for a majority of the methods described in this chapter.

INTRODUCTION TO MODELING MOLECULAR EVOLUTION

In Chapters 4 and 7, we introduced the outlines of a model for the evolution of a discrete character along the branches of a phylogeny. We stipulated that characters evolve independently and that they may switch among a finite number of character states. We also specified that traits evolve along the branches of a tree,

and that these branches have a defined (but maybe unknown) duration. These factors provide a sufficient basis for visualizing trait evolution and for seeing why a method such as parsimony might work. However, to really understand trait evolution and to apply model-based approaches such as maximum likelihood and Bayesian inference, it is necessary to add more mathematical details to this verbal model.

The core of any mathematical model of character evolution is a ***substitution model***, which specifies the way in which characters are permitted to evolve between states as well as the relative rates of different kinds of evolutionary change. All models in widespread use in phylogenetics are continuous-time ***Markov models***; that is, they describe a process in which the probability of an event happening in some time window is dependent only on the state at that time and independent of how it came to be in that state. Coin tossing is usually modeled as a Markov process. In such a model there is some *fixed* rate at which heads show up (probably 0.5 per toss). This means that the probability of heads is unaffected by the number of heads that were obtained in previous throws.

Mathematical models have now been developed for many different types of traits. We will focus on models of DNA sequence evolution. These are easy because there are only four possible states corresponding to the four bases, A, C, G, and T. While indels occur, these are not included in the basic models of sequence evolution because they add too much complexity. Instead, gap characters are treated as missing data, representing the fact that we are uncertain whether nucleotide positions that are absent should be scored as an A, C, G, or T.

We will also restrict our discussion to substitution models that are ***time reversible***. This means that the probability of a change between two states is equal in the forward and reverse directions. For example, the number of changes from A to T is assumed to be equal to the number of changes from T to A. While time reversibility is not strictly required, it provides a reasonable simplification in cases where we think that base composition (i.e., the relative frequency of the four bases, A, C, G, and T) has not changed systematically over time.

We will assume that at any one moment in time a particular position in a DNA sequence is occupied by one of the four bases and that every so often a mutation occurs. The best way to visualize this mutation is as a two-step process: the old base is removed and then a base is drawn at random from a pool of possible bases and is inserted in the position of the removed base. If the inserted base were the same as the deleted one, no change would be visible. Only if one of the three other bases were drawn from the pool would a change be visible.

We will assume that both steps, and thus the entire mutational process, occur in zero time.

To get a feel for this process, we will begin with an analogy. Imagine a card sitting face-up on the table in front of you next to a deck of cards. We will pretend that these cards do not have numbers, but are characterized just by their suit. Now imagine an invisible card-changing fairy who occasionally removes the card on the table, replaces it in the deck, and then draws a card at random from the deck (a deck with an equal number of cards of each suit) to put back onto the table. Being a fairy, this process happens faster than the human eye can track. If the new card is the same suit as the old card, there would be no visible change, but otherwise the card would change suit. You can imagine watching a card on the table and seeing it occasionally changing to a new suit. Let us now develop a mathematical description of the way the card changes.

Suppose that although the fairy changes cards at a specific rate, being flighty, he does so in an unpredictable manner, analogous to the way that radioactive atoms decay. We will focus on how frequently the fairy changes the identity of a card's suit. We will ignore events where a card is replaced by another card of the same suit (although the underlying math does deal with these "hidden" events). Let us call the rate at which suits change the substitution rate, μ.

Suppose the fairy changes cards at a rate of 0.6 substitutions per minute. This would predict that if we watched a single card for an hour, we should see an average of 36 changes (0.6×60 minutes). However, the actual number would vary hour to hour and, even more so, minute to minute. Sometimes the fairy would change the card several times in quick succession, but other times he would wait a long time between changes. However, while the waiting time between successive changes will be variable, it will average 1 minute, 40 seconds (= 1/0.6 minutes). (For the mathematically inclined, the waiting time is exponentially distributed with a mean of $1/\mu$ minutes.)

If you watch the card for a long time, you will see it visiting all four suits. A useful way to keep track of the pattern of changes is with a substitution matrix, such as that shown in Figure 8.1. This shows the expected number of substitutions of each type seen in a certain amount of time. Of the 16 possible substitutions, only the 12 off-diagonal events result in a visible change in the card. Because each of the 12 possible changes should occur equally frequently, and because the overall rate of visible change is μ, the rate at which each of the 12 changes occurs is $\mu/12$. For example, if the overall rate of change were 0.6 changes/minute, then the rate at which any of the specific kinds of change

		To:			
		♠	♦	♥	♣
From:	♠	—	$1/12\,\mu t$	$1/12\,\mu t$	$1/12\,\mu t$
	♦	$1/12\,\mu t$	—	$1/12\,\mu t$	$1/12\,\mu t$
	♥	$1/12\,\mu t$	$1/12\,\mu t$	—	$1/12\,\mu t$
	♣	$1/12\,\mu t$	$1/12\,\mu t$	$1/12\,\mu t$	—

FIGURE 8.1 Expected number of changes by the card-changing fairy in *t* minutes. The overall rate of card evolution is μ substitutions per minute.

		To:			
		♠	♦	♥	♣
From:	♠	$-\mu$	$\mu/3$	$\mu/3$	$\mu/3$
	♦	$\mu/3$	$-\mu$	$\mu/3$	$\mu/3$
	♥	$\mu/3$	$\mu/3$	$-\mu$	$\mu/3$
	♣	$\mu/3$	$\mu/3$	$\mu/3$	$-\mu$

FIGURE 8.2 Instantaneous rates of substitution by the card-changing fairy. The overall rate of card evolution is μ substitutions per minute.

occurs should be 0.6/12 = 0.05 changes per minute. Thus, if we watched a card for 1000 minutes, we would expect to see about 50 of each kind of substitution.

The substitution matrix reports the expected number of each kind of change as a function of the substitution rate, μ, and time, t. In order to make more useful predictions, we need to express this model of card "evolution" in terms of ***instantaneous rates*** of change. These are summarized in Figure 8.2. The entries in the matrix report the rate at which a card starting at the suit shown to the left will switch to each of the other suits. Since the rate of change is μ, and there are three possible alternative suits at equal frequency in the deck, the rate of change to each is $\mu/3$. The "rates" of staying in the present base are set to $-\mu$, which means that the sum of each row is zero. One way to think about this is that, because every change from the starting base results in a suit within the row, the net rate of leaving a row must be zero.

Before developing the fairy metaphor further, let us clarify the link to the goals of phylogenetic inference. Phylogenetic inference is essentially an attempt

to determine how long ago a pair of taxa last shared common ancestry. Rather than counting time in units of years, which can be done but adds complications (Chapter 11), we will focus on how far apart two taxa are from each other in substitutional units. That is to say, we attempt to determine their ***evolutionary distance***, which for DNA sequence data is the average number of substitutions that have occurred at each nucleotide position. Returning now to the fairy metaphor, we shall see why it is necessary to analyze whole sequences, comprising multiple aligned nucleotide positions rather than just one position at a time.

Suppose we put a spade card on a table and leave the room for 10 minutes. Based on the suit of the card when we return, can we say anything about the number of changes that happened while we were away (the number of changes being analogous to evolutionary distance)? To answer this we need to know how the probability of observing each of the four suits changes as a function of the substitution rate, μ, and time, t. What we want is a ***substitution probability matrix***, which shows the probability that a card starting at a particular suit would be found to be in each of the four possible suits t time units later.

A substitution probability matrix is not easy to figure out by hand because there are an infinite number of histories that start with a spade and end with, say, a heart. There could have been one change directly from a spade to a heart. But there could have been, for example, a change from spades to clubs, then a change from clubs back to spades, then a change from spades to diamonds, and finally a change from diamonds to hearts. Thankfully, with the help of calculus and matrix algebra, substitution probability matrices can be derived from instantaneous rate matrices. The substitution probability matrix for the fairy model is given in Figure 8.3. The base of natural logarithms, e, emerges when we mathematically integrate over the alternative possible histories.

		To:			
		♠	♦	♥	♣
From:	♠	$1/4 + 3/4e^{-4/3\mu t}$	$1/4 - 1/4e^{-4/3\mu t}$	$1/4 - 1/4e^{-4/3\mu t}$	$1/4 - 1/4e^{-4/3\mu t}$
	♦	$1/4 - 1/4e^{-4/3\mu t}$	$1/4 + 3/4e^{-4/3\mu t}$	$1/4 - 1/4e^{-4/3\mu t}$	$1/4 - 1/4e^{-4/3\mu t}$
	♥	$1/4 - 1/4e^{-4/3\mu t}$	$1/4 - 1/4e^{-4/3\mu t}$	$1/4 + 3/4e^{-4/3\mu t}$	$1/4 - 1/4e^{-4/3\mu t}$
	♣	$1/4 - 1/4e^{-4/3\mu t}$	$1/4 - 1/4e^{-4/3\mu t}$	$1/4 - 1/4e^{-4/3\mu t}$	$1/4 + 3/4e^{-4/3\mu t}$

FIGURE 8.3 **Substitution probability matrix under a simple model of card "evolution" with equal frequencies of the four suits.** The mutation rate, in card changes per minute, is denoted μ. The number of minutes over which evolution is allowed to happen is denoted t.

Before moving on, it is worth looking briefly at the matrix and trying to develop a feel for what it communicates. What is the probability of starting as a spade and ending as a spade some very short period of time later? If μt is small, then $e^{-\mu t}$ is close to 1.0 (remember, any number taken to the power of 0 has a value of 1). In that case, the probability of still being a spade is close to ¼ + ¾, or 1.0, and the probability of being in any other state is almost zero. This makes sense. If no time has passed, you must still be in the state you started in.

Now consider the other extreme, when μt is very large. In that case $e^{-\mu t}$ is very close to zero, meaning that the probability of being in any of the four states is ¼, or 0.25. This too makes sense. If you wait an infinite amount of time, the ending suit is no longer constrained by the starting suit. In that case, with all the suits occurring at equal frequency in the deck, there is a 25% chance that the card now showing is of each of the four suits. For any time between 0 and ∞, the probability of the card still being a spade is between 0.25 and 1.0, while the probability of the card being another suit is between 0 and 0.75. Table 8.1 lists some probabilities for different values of μt.

Now we can return to the original question: If we leave the card for 10 minutes, can we estimate the number of changes of suit that occurred while we were out of the room? If we knew the substitution rate, μ, then we would *know* the evolutionary distance, which is defined as μt, without even looking at the card when we return. If we did not know μ, then we would have a problem. The card is either the same or a different suit, but in either case there could have been few

TABLE 8.1 The probability of a card starting as a spade and being a spade or another suit after an average of μt substitutions have occurred (Probability of not being a spade = 1 − Probability of being a spade)

μt	Prob[♠]	Prob[not ♠]
0.01	0.990	0.010
0.05	0.952	0.048
0.1	0.906	0.094
0.5	0.635	0.365
1	0.448	0.552
5	0.251	0.749
10	0.250	0.750

(0 or 1) or infinitely many changes while we were out of the room, depending upon the value of μ. We cannot estimate μ by seeing if one card has changed state between two moments of observation.

The solution is to leave behind not one card but a whole line of cards. By looking at the proportion of the cards that had changed suit in 10 minutes, we can obtain information about μ, and this, in turn, allows us to estimate the number of changes of suit that occurred when we were not looking. Let us walk through this numerically.

Suppose we had laid out 100 cards in a row before leaving the room and noted their suit. If we came back and 60 cards had the same suit as before we left and 40 had changed, how many changes on average did each card experience? You might be tempted to say 0.4 changes per card (40/100), but this ignores all the substitutions that were then "covered up" by further changes. Using the substitution probability matrix, we can take account of the extra changes.

Since the proportion of cards that is unchanged is 0.6, our best estimate is that the probability of not changing suit is also 0.6. The substitution probability matrix shows us that we just need to find the value of μt such that $\frac{1}{4} + \frac{3}{4}e^{-4/3\mu t} = 0.6$. Some simple algebra solves the equation and finds that the evolutionary distance, μt, is 0.572. In this case, because we know that t is 10 minutes, we can also calculate the value of μ (0.57/10 = 0.057 substitutions/minute). Figure 8.4 shows how the frequency of cards that *have* changed suit

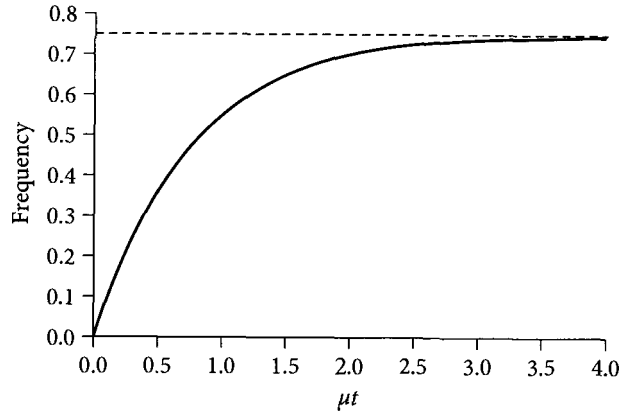


FIGURE 8.4 Expected frequency of cards that have a different suit from the ancestor as a function of μt.

changes as a function of μt. You will see that a frequency of 0.4 corresponds to $\mu t = 0.57$. This shows that if we are willing to assume that all the cards are acted upon identically by the fairy, then by looking at the proportion of suits that change between two moments of observation, we can estimate the number of hidden substitutions that happened and thereby calculate the card analog of evolutionary distance: the average number of changes of suit per card.

The card-changing fairy metaphor can now be related back to DNA sequence evolution. The four suits correspond to the four bases, A, C, G, and T. An individual card corresponds to a single position in a DNA sequence, and a line of cards corresponds to a sequence of DNA. The simple model illustrated with the fairy was originally developed by Jukes and Cantor and is usually called the Jukes-Cantor or JC model of molecular evolution (Jukes and Cantor 1969). This model assumes that (a) all four bases occur at equal frequency, (b) each kind of substitution (A to C, A to T, etc.) occurs at an equal rate, and (c) the rate of substitution is the same for all nucleotide positions in the sequence being studied. Under these assumptions, the substitution probability matrix for nucleotide positions and bases resembles that for cards and suits, as shown in Figure 8.5.

Let us think about what would happen if you let an ancestral sequence evolve while you kept track of the proportion of positions at which the sequence differs from its ancestor. At the beginning, the sequence will be identical to the ancestor, but differences will gradually accumulate. The initial rate of increase will be equal to μt because each change adds to the difference between the ancestor and descendant. However, as time continues, more and more changes will fail to generate additional differences. For example, if the ancestor had a G, which subsequently changed to a C, a later change to a T at this position would not be

		To:			
		A	C	G	T
From:	A	$1/4 + 3/4e^{-4/3\mu t}$	$1/4 - 1/4e^{-4/3\mu t}$	$1/4 - 1/4e^{-4/3\mu t}$	$1/4 - 1/4e^{-4/3\mu t}$
	C	$1/4 - 1/4e^{-4/3\mu t}$	$1/4 + 3/4e^{-4/3\mu t}$	$1/4 - 1/4e^{-4/3\mu t}$	$1/4 - 1/4e^{-4/3\mu t}$
	G	$1/4 - 1/4e^{-4/3\mu t}$	$1/4 - 1/4e^{-4/3\mu t}$	$1/4 + 3/4e^{-4/3\mu t}$	$1/4 - 1/4e^{-4/3\mu t}$
	T	$1/4 - 1/4e^{-4/3\mu t}$	$1/4 - 1/4e^{-4/3\mu t}$	$1/4 - 1/4e^{-4/3\mu t}$	$1/4 + 3/4e^{-4/3\mu t}$

FIGURE 8.5 Substitution probability matrix under the JC model of DNA sequence evolution. The mutation rate, in substitutions per unit time, is denoted μ. The time interval over which evolution is allowed to happen is denoted t.

a new difference. Furthermore, some changes can actually reduce the difference by bringing the descendant back to the state found in the ancestor. Thus, over time the rate of increase of distance will slow down until eventually the distance is "stuck" at 0.75 (see Figure 8.4).

MORE REALISTIC MODELS OF MOLECULAR EVOLUTION

While the JC model provides an accessible starting point for thinking about models of molecular evolution, all three of its core assumptions are violated by real DNA sequences. The four bases are usually not present at equal frequencies, some kinds of substitutions occur at different rates than others, and some positions in a DNA sequence have a higher rate of evolution than others. Over the last four decades, more sophisticated models have been developed that more accurately reflect these and other violations of the JC model.

The first extension of the JC model is to allow bases to occur at different frequencies. This model is usually called F81, because of when and by whom it was proposed (Felsenstein 1981a). It is analogous to assuming that the deck from which the fairy draws the card does not necessarily have an equal proportion of the four suits. There are several possible reasons why base composition is expected to be uneven. For example, selection can favor a higher frequency of guanine and cytosine in some RNA molecules (e.g., ribosomal RNAs) because these two bases form three rather than two hydrogen bonds, resulting in a more stable secondary structure. Whatever the cause, uneven base composition affects both the probability that a site will start in a certain base and the probability that it will be replaced by each of the other bases. Let us go through this more fully to illustrate the ways in which a basic model can be extended to include more biological realism.

Let us use π to represent the frequency of each base in the "pool" (analogous to the deck of cards) from which bases are drawn. Since there are only four bases, the base frequencies (π_A, π_C, π_G, π_T) add up to 1.0. Assuming that a sequence starts in equilibrium (where the frequency of bases in the sequence matches the frequency in the pool), the probability of a position being in state G at time 0 is equal to π_G. After a mutation has happened at a site, the probability that the new base will be a G is, again, π_G. This means that the expected number of mutations of a given type is μt times the product of the frequencies of the starting and ending bases. The matrix in Figure 8.6 shows the expected

		To:			
		A (freq = π_A)	C (freq = π_C)	G (freq = π_G)	T (freq = π_T)
From:	A (freq = π_A)	—	$\pi_A\pi_C\mu t$	$\pi_A\pi_G\mu t$	$\pi_A\pi_T\mu t$
	C (freq = π_C)	$\pi_C\pi_A\mu t$	—	$\pi_C\pi_G\mu t$	$\pi_C\pi_T\mu t$
	G (freq = π_G)	$\pi_G\pi_A\mu t$	$\pi_G\pi_C\mu t$	—	$\pi_G\pi_T\mu t$
	T (freq = π_T)	$\pi_T\pi_A\mu t$	$\pi_T\pi_C\mu t$	$\pi_T\pi_G\mu t$	—

FIGURE 8.6 Expected numbers of each type of substitution under the F81 model of DNA sequence evolution. The frequency of each base (A, C, G, and T) is indicated with the subscripted notation π, where ($\pi_A + \pi_C + \pi_G + \pi_T = 1$).

		To:			
		A (freq = π_A)	C (freq = π_C)	G (freq = π_G)	T (freq = π_T)
From:	A (freq = π_A)	$-m(\pi_C + \pi_G + \pi_T)$	$\pi_C m$	$\pi_G m$	$\pi_T m$
	C (freq = π_C)	$\pi_A m$	$-m(\pi_A + \pi_G + \pi_T)$	$\pi_G m$	$\pi_T m$
	G (freq = π_G)	$\pi_A m$	$\pi_C m$	$-m(\pi_A + \pi_C + \pi_T)$	$\pi_T m$
	T (freq = π_T)	$\pi_A m$	$\pi_C m$	$\pi_G m$	$-m(\pi_A + \pi_C + \pi_G)$

FIGURE 8.7 The instantaneous rate matrix under the F81 model of DNA sequence evolution. Base frequency notation is the same as Figure 8.6. The effective mutation rate, after correcting for base compositional inequality (see text), is denoted m.

frequency of all 12 kinds of change. If you have studied genetics, you will note a resemblance here to the Punnett square method of predicting genotype frequencies. You will also observe that the expected number of A to G changes, $\pi_A\pi_G\mu t$, is the same as the reverse, $\pi_G\pi_A\mu t$. This means that, at equilibrium, the base frequency will tend to remain unchanged. It also means that the F81 model (like the Jukes-Cantor model) is time reversible: evolution looks the same whether it runs forward or backward in time.

The F81 instantaneous rate matrix (Figure 8.7) is derived similarly to the Jukes-Cantor model. Because the frequency of the starting base does not matter (the matrix shows the rate of a substitution conditioned on the identity of the starting base), the rates are influenced only by the frequency of the ending base. The diagonals are again such that the rows add up to zero. To be strict, it should be noted that the m included in this matrix is a modified version of μ used in the JC model. This modification is needed to account for the effect

of base frequencies on the instantaneous rate of substitution. Rare bases will tend to persist for less time (manifested as a higher rate of substitution) than common bases because common bases will often be substituted by themselves, resulting in no actual change. While it is not necessary to know the derivation of m to understand the principles, we provide the formula for completeness: $m = \mu(1 - \pi_A^2 - \pi_C^2 - \pi_G^2 - \pi_T^2)$.

The substitution probability matrix for the F81 model is shown in Figure 8.8. You will see that the probability of a substitution in each direction, for example, C to G versus G to C, can be different because the probabilities depend on the frequencies of the bases. This may seem to be at odds with Figure 8.6, which shows time reversibility (the expected number of substitutions in each direction is the same). The apparent discrepancy is explained by the fact that when a position is occupied by a rare base, it will tend, on average, to quickly switch to another state, but when a common base is present, it will tend to persist longer. This difference in the waiting time results from the fact that when a mutation event happens, it is relatively probable that the common base will be replaced by itself, resulting in no actual substitution.

For the mathematically inclined, you may notice that if $\pi_A = \pi_C = \pi_G = \pi_T = 0.25$, then $m = 3\mu/4$ or $\mu = 4m/3$. If you plug this into the diagonal entries, they become $\frac{1}{4} + \frac{3}{4}e^{-4/3\mu t}$ and the off-diagonal values become $\frac{1}{4} - \frac{1}{4}e^{-4/3\mu t}$. These are identical to the JC model, showing that JC is a special case of F81 in which the four bases are at equal frequency.

The next level of complexity that molecular models consider involves allowing for different relative rates for different kinds of substitution. For example, suppose that the fairy had a propensity to replace a card by a suit of the same

		To:			
		A	C	G	T
From:	A	$\pi_A + (1 - \pi_A)e^{-mt}$	$\pi_C(1 - e^{-mt})$	$\pi_G(1 - e^{-mt})$	$\pi_T(1 - e^{-mt})$
	C	$\pi_A(1 - e^{-mt})$	$\pi_C + (1 - \pi_C)e^{-mt}$	$\pi_G(1 - e^{-mt})$	$\pi_T(1 - e^{-mt})$
	G	$\pi_A(1 - e^{-mt})$	$\pi_C(1 - e^{-mt})$	$\pi_G + (1 - \pi_G)e^{-mt}$	$\pi_T(1 - e^{-mt})$
	T	$\pi_A(1 - e^{-mt})$	$\pi_C(1 - e^{-mt})$	$\pi_G(1 - e^{-mt})$	$\pi_T + (1 - \pi_T)e^{-mt}$

FIGURE 8.8 Substitution probability matrix under the F81 model of DNA sequence evolution. Base frequency notation is the same as Figure 8.6. The effective mutation rate, after correcting for base compositional inequality (see text), is denoted m. The time interval over which evolution is allowed to happen is denoted t.

color (even beyond any bias that might arise based on the frequency of the four suits). This would mean that red-to-red and black-to-black substitutions would tend to happen at a higher rate than red-to-black or black-to-red substitutions. This turns out to be very similar to what happens biochemically to DNA during the mutational process because of the structural differences between purine (A and G) and pyrimidine (C and T) bases. As a result, transitions (purine-to-purine or pyrimidine-to-pyrimidine substitutions) tend to happen at a higher rate than transversions (purine-to-pyrimidine or pyrimidine-to-purine substitutions). This phenomenon is termed ***transition:transversion bias.***

The method for accommodating this inequality (or any other) in substitution rate is to include a parameter in the model that adjusts the rate of one class of change relative to the other(s). In the case of the HKY (Hasegawa, Kishino, and Yano 1985) model, which includes both unequal base frequencies and transition:transversion bias, the relative rate parameter, denoted κ, is added to the rate matrix as a multiplier to the transitions (Figure 8.9): the higher the value of κ, the higher the rate of transitions relative to transversions.

This idea can be extended by allowing the two different kinds of transitions and/or the four different kinds of transversions to have different rate modifiers. This is achieved by adding parameters to the rate matrix that indicate the rate of certain substitutions relative to others. The most extreme case that is commonly used is the general time-reversible model, or GTR. This adds rate multipliers to five of the six rates of change (the six off-diagonal elements in the matrix). The sixth change does not need a multiplier because it is implicitly set to a rate of 1.0. Being time reversible, the changes in both directions (e.g., A to G and G to A) use the same rate multiplier.

		To:			
		A (freq = π_A)	C (freq = π_C)	G (freq = π_G)	T (freq = π_T)
From:	A (freq = π_A)	$-m(\pi_C + \kappa\pi_G + \pi_T)$	$\pi_C m$	$\pi_G \kappa m$	$\pi_T m$
	C (freq = π_C)	$\pi_A m$	$-m(\pi_A + \pi_G + \kappa\pi_T)$	$\pi_G m$	$\pi_T \kappa m$
	G (freq = π_G)	$\pi_A \kappa m$	$\pi_C m$	$-m(\kappa\pi_A + \pi_C + \pi_T)$	$\pi_T m$
	T (freq = π_T)	$\pi_A m$	$\pi_C \kappa m$	$\pi_G m$	$-m(\pi_A + \kappa\pi_C + \pi_G)$

FIGURE 8.9 The instantaneous rate matrix under the HKY model of DNA sequence evolution. Notation is the same as Figure 8.7 except for the addition of a rate multiplier, κ, which indicates how many times faster transitions occur than transversions.

The GTR model includes more free parameters than any of the other models we have described. Nonetheless, GTR does not require that there be an actual deviation from equal base frequencies, nor that the five substitution types be different from one another. The JC, F81, and HKY are all special cases of the GTR model or, put another way, they are nested within the GTR model. The HKY model is derived from GTR when the four transversions are equal and the two transitions are equal. F81 is derived when, in addition, the rates of transversions equal the rates of transitions. And, finally, JC is derived when, in addition, the base frequencies are equal. The connections among these models are shown in Figure 8.10.

The final major assumption of the models discussed so far is that all sites in a DNA sequence evolve at the same rate. However, we commonly expect

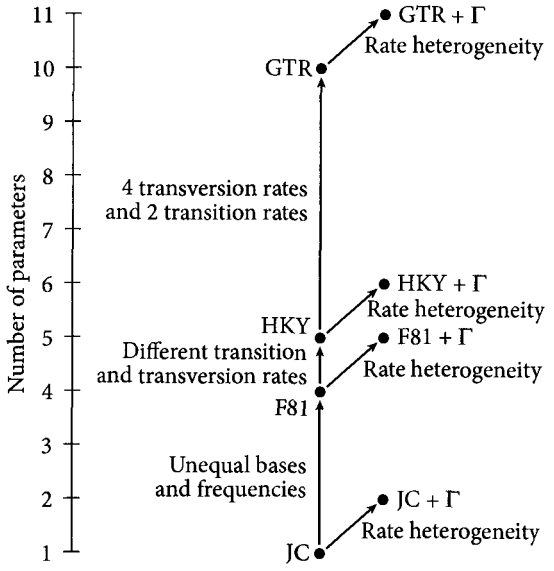


FIGURE 8.10 Depiction of the relationship between some commonly used models of evolution. Any two models that are connected by arrows that proceed in the same direction are nested: the simpler model (closer to the bottom of the chart) contains a subset of the free parameters of the more complex model. The axis on the left shows the number of free rate parameters in the model. The figure assumes that site-to-site rate heterogeneity is modeled using a discrete approximation to a gamma (Γ) distribution, which adds one free parameter to the model.

variation in rates among sites. This is expected given that selection will tend to slow down the rate of evolution of positions that critically affect gene function, which results in a higher rate of extinction of alleles that have variants at those sites. But how can we allow for rate variation while still being able to use the frequency of unchanged positions to infer μt?

One approach to deal with rate inequality would be to use prior information to divide a DNA sequence into ***partitions***, sets of positions, each with the same rate of molecular evolution. For example, we could apply the JC model separately to subsets of the sites that we believe have the same elevated or depressed rate of evolution. The problem with this is that it requires that, before looking at the data, we predict which subsets of sites will have the same rate of evolution. This is difficult to do in most cases.

The alternative is to assume that the rate of substitution, μ, is not the same for all sites, but is drawn from a distribution of rates. By defining a form for the distribution of rates among sites (most commonly a gamma distribution), and by assigning different sites to different rate categories, it is possible to allow for site-to-site rate heterogeneity without having to decide, in advance, which sites are rapidly or slowly evolving. The details are covered by some of the recommended further readings.

In addition to rate heterogeneity, models of molecular evolution can allow for the possibility of nonindependence between nucleotide positions. This might arise, for example, due to base pairing during folding of RNA molecules or the translation of the three bases of each codon into an amino acid. Also, as elaborated in Chapter 11, we can use molecular clock models, which use the same basic substitution models but place constraints on the lengths of branches to force all living species to be the same evolutionary distance from the root.

As the foregoing illustrates, it is possible to develop more and more sophisticated models of evolution to accommodate our knowledge of how DNA sequences actually evolve. The details of these more sophisticated models are less important to grasp than the general principles: we can build realistic models of how DNA sequences evolve and use them to calculate the probability that particular substitutions occur. We can also use the data to guide the selection of an appropriate model of evolution, as discussed later in this chapter.

Up until now we have only considered models of DNA evolution, but phylogeneticists employ many other kinds of data. For many, but not all, of the classes of data described in Appendix 1, continuous-time Markov models have been developed. These include protein sequences, morphology, indels,

restriction fragment length polymorphisms, and amplified fragment length polymorphisms. As a result, model-based methods of phylogenetic inference are now available for almost all widely used data types.

DISTANCE METHODS

The core principle underlying distance methods is that if we knew the true evolutionary distances between each pair of taxa (defined as the average number of substitutions per site in a DNA sequence), then these distances would correspond to only one tree. Because the evolutionary distance between any two taxa is the sum of the lengths (evolutionary distances) of all of the branches on the path between those two taxa, knowing the true evolutionary distances amounts to knowing the tree. For example, if the tree shown in Figure 8.11 were correct, the true evolutionary distances between each pair of taxa would be those given in Table 8.2. If you were given just the distances in Table 8.2, you could work backward to draw an unrooted version of the phylogram in Figure 8.11. The

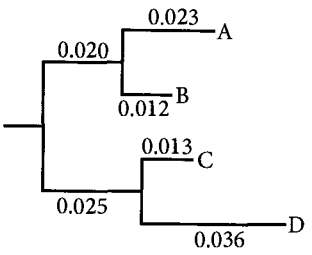


FIGURE 8.11 Phylogram showing evolutionary distances (in substitutions/site).

TABLE 8.2 Evolutionary distances between the taxa in Figure 8.11

	A	B	C	D
A	0			
B	0.035	0		
C	0.081	0.070	0	
D	0.104	0.093	0.049	0

aim of distance methods is to determine the evolutionary distances between taxa and then use those to infer the true phylogeny.

The starting point for distance-based phylogenetic methods is usually the calculation of the proportion of traits that differ between those taxa, their ***pairwise distance.*** For this example, Table 8.3 lists the first 10 characters from the carnivoran morphology data set (from Tables 7.2 and 7.4). The pairwise distance is the proportion of characters for which a pair of taxa have a different character state. For example, the outgroup and the cat differ at five characters (2, 4, 7, 8, and 9). We divide this by the total number of characters to get their pairwise distance, 5/10, or 0.5. The pairwise distance matrix for these taxa based on the data in Table 8.3 is shown in Table 8.4.

Table 8.5 provides the pairwise distances using the carnivoran molecular data. This is calculated the same as for morphological data. However, the convention is to only count sites where both taxa have the character scored. Thus, positions that are missing in one or both taxa (including indels) are excluded in the calculation of pairwise distance.

The simple-minded approach to phylogenetic inference would be to take these observed distances and suppose that they are reasonable estimates of the

TABLE 8.3 Ten characters from the carnivoran morphology data set

	Characters									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Outgroup	0	0	0	0	0	0	0	0	0	0
Cat	0	1	0	1	0	0	1	1	1	0
Hyena	0	1	0	1	0	0	1	0	1	0
Civet	0	1	0	0	0	0	0	0	1	0
Dog	1	0	0	0	1	0	0	0	0	0
Raccoon	1	0	0	0	1	0	0	0	0	0
Bear	1	0	0	0	1	1	0	0	0	1
Otter	1	0	0	0	1	0	0	0	0	1
Seal	1	0	1	0	1	1	0	0	0	1
Walrus	1	0	1	0	1	1	0	0	0	1
Sea lion	1	0	1	0	1	1	0	0	0	1

TABLE 8.4 Pairwise distances for the morphological data

	Outgroup	Cat	Hyena	Civet	Dog	Raccoon	Bear	Otter	Seal	Walrus
Cat	0.5									
Hyena	0.4	0.1								
Civet	0.2	0.3	0.2							
Dog	0.2	0.7	0.6	0.4						
Raccoon	0.2	0.7	0.6	0.4	0					
Bear	0.4	0.9	0.8	0.6	0.2	0.2				
Otter	0.3	0.8	0.7	0.5	0.1	0.1	0.1			
Seal	0.5	1	1	0.7	0.3	0.3	0.1	0.2		
Walrus	0.5	1	1	0.7	0.3	0.3	0.1	0.2	0	
Sea lion	0.5	1	1	0.7	0.3	0.3	0.1	0.2	0	0

TABLE 8.5 Pairwise distances for the carnivoran molecular data

	Mole	Cat	Hyena	Civet	Dog	Raccoon	Bear	Otter	Seal	Walrus
Cat	0.244									
Hyena	0.269	0.092								
Civet	0.246	0.081	0.092							
Dog	0.277	0.190	0.205	0.198						
Raccoon	0.288	0.206	0.212	0.204	0.190					
Bear	0.270	0.175	0.185	0.171	0.179	0.135				
Otter	0.296	0.205	0.209	0.199	0.207	0.121	0.148			
Seal	0.289	0.190	0.196	0.189	0.182	0.151	0.126	0.154		
Walrus	0.283	0.187	0.199	0.183	0.189	0.146	0.128	0.147	0.056	
Sea lion	0.287	0.193	0.198	0.187	0.188	0.147	0.127	0.149	0.058	0.028

true evolutionary distances. Then we could search for a tree that comes closest to predicting this set of evolutionary distances. While it is not uncommon to use this approach, it is inadvisable. As discussed in the context of the card-changing fairy, pairwise distances accumulate more slowly than evolutionary distances because multiple changes occurring at the same site do not always increase the pairwise distance. As a result, the first step in distance analysis is generally to estimate evolutionary distances from observed pairwise distances by correcting the distances based on expectations calculated under a particular model of character evolution.

A pairwise distance can be converted to an evolutionary distance by using the expected relationship between pairwise distance and evolutionary distance. Figure 8.12 shows this for the Jukes-Cantor model of DNA sequence evolution. This graph is the same as Figure 8.4 except that the axes have been relabeled to fit the current context (distance between two taxa rather than frequency of differences between an ancestor and a descendant). Given such a graph, a pairwise distance of 0.24 would be converted into an evolutionary distance of 0.3. Table 8.6 gives the estimated evolutionary distances for the carnivoran molecular data under the Jukes-Cantor model.

Having estimated evolutionary distances, distance methods aim to find the tree that is most consistent with these distances. The first approach is to conduct a series of calculations on the distance matrix that lead directly to a tree. The most widely used method is the ***neighbor-joining*** (NJ) algorithm (which is described in detail in Swofford et al. 1996 and many online resources). NJ has

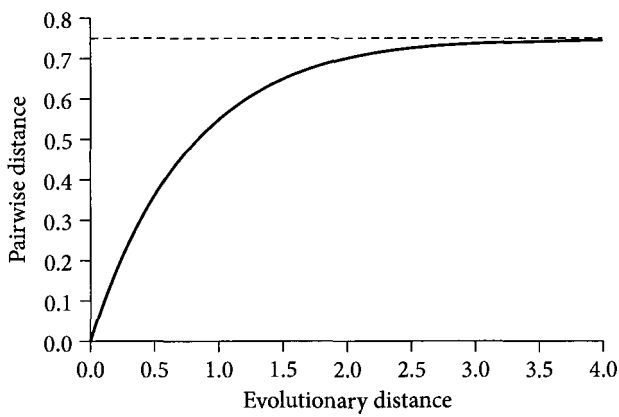


FIGURE 8.12 Relationship of evolutionary and pairwise distances under the JC model. The dashed line indicates the maximum expected pairwise distance, 0.75.

TABLE 8.6 Evolutionary distances estimated under the Jukes-Cantor model for the carnivoran molecular data

	Mole	Cat	Hyena	Civet	Dog	Raccoon	Bear	Otter	Seal	Walrus
Cat	0.300									
Hyena	0.343	0.098								
Civet	0.305	0.086	0.099							
Dog	0.362	0.225	0.244	0.236						
Raccoon	0.371	0.243	0.251	0.239	0.223					
Bear	0.345	0.201	0.215	0.197	0.208	0.150				
Otter	0.389	0.241	0.247	0.233	0.247	0.134	0.166			
Seal	0.376	0.223	0.231	0.222	0.213	0.171	0.140	0.175		
Walrus	0.366	0.218	0.234	0.214	0.222	0.164	0.142	0.166	0.058	
Sea lion	0.373	0.226	0.234	0.219	0.220	0.166	0.140	0.168	0.061	0.029

the virtue that, if the corrected pairwise distances correspond exactly to a tree that would predict those distances, then this tree will be identified by the algorithm. It is also an extremely quick method: even for a very large data set, a tree can be obtained in a fraction of a second. The downside of NJ is that it yields a tree but does not attach a measure of quality to that tree. This means that NJ does not allow us to determine whether a particular tree is significantly better or worse than another. For these reasons, NJ is widely used in situations where we need a quick, but approximate estimate of the true tree. Figure 8.13 shows the neighbor-joining tree for the carnivoran molecular data obtained using the JC distances.

Most other methods for estimating trees from a distance matrix use optimality criteria, that is, measures of tree quality. In parsimony, the optimality criterion is tree length and we search for the tree with the lowest length. For distance methods, a widely used optimality approach is to search for the tree (with branch lengths) that minimizes the difference between the distances in the matrix and the distances that are predicted by the tree. The first step is to calculate the expected evolutionary distances between each pair of taxa on a tree by summing up all of the intervening branch lengths. The set of expected distances is compared to the corrected pairwise distances. One way to quantify

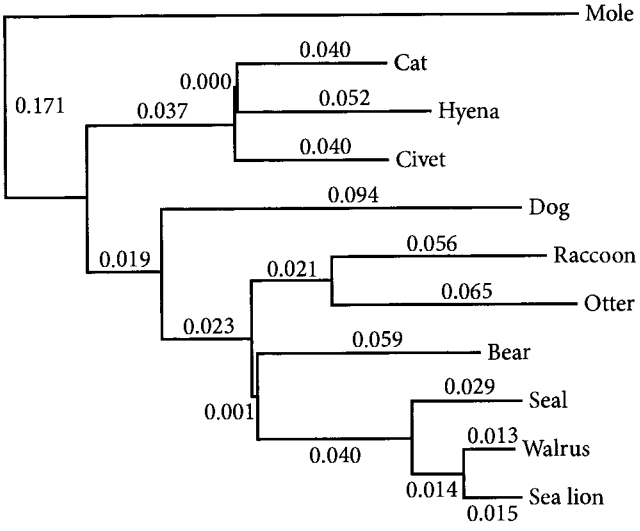


FIGURE 8.13 Neighbor-joining tree based on the JC distances obtained from the carnivoran molecular data. Branch lengths are given in average number of substitutions per site (evolutionary distance).

the fit between a tree and the observed distances is to sum the squared differences between each observed and expected distance. We can then search for a tree topology and a set of branch lengths that minimizes this metric. This tree could be said to be optimal because it comes closest to predicting the observed distances.

There are several variants of this basic approach (see Further Reading). The one we will discuss here is ***minimum evolution***. This starts by choosing the optimal branch lengths for a given tree topology using the least squares method: adjusting branch lengths to minimize the sum of the squared deviations between the observed and expected distances. However, rather than picking the tree topology that minimizes this same metric, minimum evolution picks the tree topology on which the total branch length (the sum of the lengths of all the branches) is minimized.

A heuristic search can be conducted to find the optimal tree for a particular distance matrix. These are similar to the searches described for parsimony (Chapter 7) except that the value that is calculated for each tree that is visited is not the parsimony score but the measure of fit between the tree and the distance data (either the squared deviation or the sum of the lengths of all the branches). In comparison to parsimony, which only considers topology when determining the score of a tree, a distance search needs to explore branch length. This explains why distance optimality methods tend to be significantly slower than parsimony for the same number of taxa.

Figure 8.14 shows the distance tree estimated from the carnivoran molecular data using the minimum evolution method. It is worth noting that this tree does not exactly predict the observed distances. For instance, the branches between mole and cat sum to 0.251 (0.172 + 0.037 + 0.001 + 0.041) although their estimated evolutionary distance was 0.300 (Table 8.6). This difference could be due to chance events during evolution, errors in the estimation of evolutionary distances from observed pairwise distance, or both. This tree is similar but not identical to the NJ tree, which is expected given that they both used the same evolutionary distance matrix.

When converting character state data to distances, some information is discarded. There is only one distance matrix for a given character state matrix, but many character state matrices can yield the same distance matrix. Because of this loss of information, distance methods tend to have less statistical power than character state matrix approaches, such as parsimony or maximum likelihood.

Minimum evolution and similar optimality methods remain useful in cases in which the original data are already in the form of a pairwise distance (e.g.,

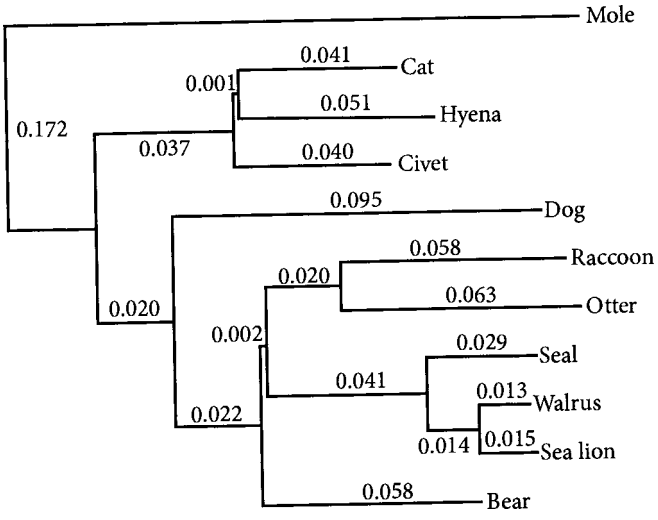


FIGURE 8.14 Minimum evolution tree based on the JC distances obtained from the carnivoran molecular data. Branch lengths are given in average number of character state changes per character.

DNA–DNA hybridization data). Also, it has been found that some specialized distance corrections provide computationally efficient ways to deal with cases where base composition varies among taxa. Nonetheless, you will encounter distance optimality methods much less frequently than neighbor-joining, maximum likelihood, or Bayesian methods.

MAXIMUM LIKELIHOOD

The maximum likelihood (ML) criterion is not specific to phylogenetics, but is a general approach used throughout statistics. Indeed, the early development of parsimony was guided by a desire that it should approximate maximum likelihood. However, the computational complexity of ML delayed its implementation as a method of phylogenetic analysis until the 1990s.

The application of ML to phylogenetics involves searching for the tree that has the highest probability of giving rise to the observed data. Before delving into the application of likelihood to biological data, let us begin by exploring the underlying principles using a coin example.

Suppose you have a bag of coins and you know that half of the coins are fair (50% chance of a head) and half of the coins are biased (75% chance of a head). You draw one coin from the bag and wish to consider two alternative hypotheses: the coin is fair versus the coin is biased. The coin is tossed 10 times and each time it falls heads-up. You may now apply likelihood to ask whether the observed data (10 heads) support one of the hypotheses and, furthermore, whether the data are decisive enough to make it reasonable to reject the alternative hypothesis.

The first stage is to specify a model of how coin tossing works. Let us assume that coins all have a head on one side and a tail on the other, that each toss is independent of previous tosses, and that you are 100% accurate in distinguishing heads and tails. Under this model we can calculate the ***likelihood***, which is defined as the probability that the data would have arisen under the hypothesis. With 10 heads, the likelihood for a fair coin is 0.5^{10} or ~0.00098 (Figure 8.15). This result does *not* mean that there is a 0.1% chance that the coin is fair. It just means that there is a 0.1% chance of this specific outcome for a fair coin.

How probable are the observed data under the hypothesis that the coin is biased? The likelihood under the bias hypothesis is 0.75^{10} or ~0.0563 (Figure 8.15). This is still a low number, which tells us that even under the bias hypothesis this particular outcome is improbable. What counts, however, is not the absolute value of the likelihood but a comparison of the likelihoods of the two competing hypotheses. The data are 0.0563/0.00098, or approximately 56 times as probable under the biased coin hypothesis as under the fair coin hypothesis. This ***likelihood ratio***, which is usually presented as a natural logarithm, is a measure of the evidential support for one hypothesis over the other. In this case, the log-likelihood ratio is ln(56) = 4.02. Because 4.02 is well above 2.0, a

Toss	1	2	3	4	5	6	7	8	9	10	Likelihood
Result											
Prob. if fair	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.001
Prob. if biased	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.056

FIGURE 8.15 Likelihood of ten sequential heads for a fair or a biased coin. The likelihood is the product of the probabilities of each individual toss, e.g. 0.5^{10}, under the fair coin hypothesis.

commonly used threshold of significance (in some circumstances, a likelihood ratio of 2.0 approximates the traditional $P < 0.05$ confidence threshold), we would say that the data strongly support the conclusion that the coin is biased.

Now let us consider how likelihood is applied to phylogenetic inference. In this case, the observed data are the characters for each taxon (the character state matrix) and the hypotheses are all the possible trees. Our aim is to determine the probability of the data arising under each tree on the principle that the tree that has the highest likelihood is the best estimate of the true tree. Similar to distance methods, when we talk of a "tree" in a likelihood context, we are thinking of both topology and branch lengths because both factor into our assessment of tree quality.

The first thing we need is a mathematical model of character evolution analogous to the model that specified that the probability of a fair coin coming up heads is 0.5. To illustrate the principles, let us consider the JC model and the simplest possible tree, comprising just two taxa, from which we have obtained a six base-pair DNA sequence (Figure 8.16). The only aspect of this tree that is unknown is the length of the branch separating these two taxa. For four of the positions, both taxa have the same base, indicating that either no change happened or there was a change to a new state and back again. The fact that four out of six bases are identical provides evidence that the branch is not infinitely long because, if it were, we would expect an average of 1.5 matching bases rather than 4.0 (because, under the JC model, ¾ of the six bases should differ after an infinite amount of time; Figure 8.12). For two of the positions the two taxa have different bases, showing that the intervening branch is not zero length. The maximum likelihood criterion proposes that the best estimate of the branch length is that which yields the highest likelihood.

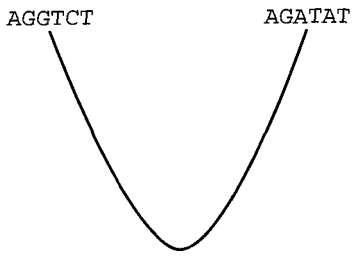


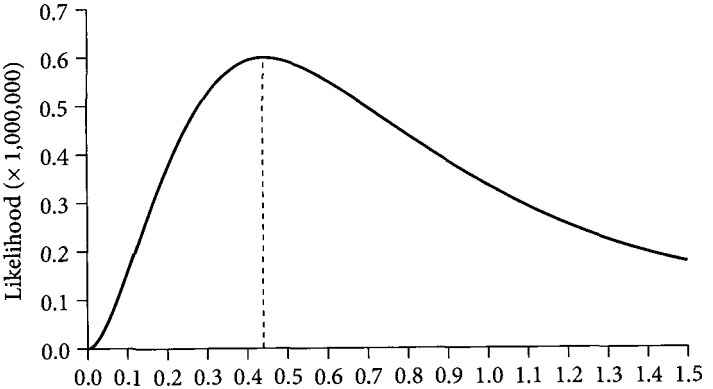
FIGURE 8.16 A two-taxon tree with six base-pair DNA sequences marked at the tips. The only element of uncertainty in this tree is the length of the single branch.

Recall that the branch length is equal to the evolutionary distance between these two taxa (the number of substitutions per site). This will be determined by the branch's duration and its rate of evolution. These two quantities are hard to disentangle (as will be discussed in the molecular dating section of Chapter 11), but all we need to worry about here is their product: μt. The relationship between branch length and probability of change is given in Figure 8.5.

In both taxa, the first base in the sequence is an A. Considering only one taxon for a moment, the probability that this first base is an A is ¼ (since, under Jukes-Cantor, we assume that the bases are at equal frequency). Given that it was an A in the first taxon, the probability that it is an A after μt units is $\frac{1}{4} + \frac{3}{4}e^{-4/3\mu t}$ (see Figure 8.5). So the probability of seeing the data at site 1 is $\frac{1}{4}(\frac{1}{4} + \frac{3}{4}e^{-4/3\mu t})$. We can substitute any value of μt (the branch length) into this equation to obtain the probability that this site would have evolved given this branch length. This probability is called the ***site likelihood***. The second site has the same pattern and thus will have the same site likelihood for any branch length. The third site has a G in one taxon (probability ¼) and an A in the other taxon. The probability of this substitution is $\frac{1}{4} - \frac{1}{4}e^{-4/3\mu t}$, giving character three a site likelihood of $\frac{1}{4}(\frac{1}{4} - \frac{1}{4}e^{-4/3\mu t})$.

The likelihood of a tree knowing the entire character matrix is given by the product of the individual site likelihoods. The likelihood is based on a product, rather than a sum, because all of the characters need to attain their observed states in order to obtain the full data matrix. When we multiply the individual site likelihoods, we obtain an overall likelihood of $[\frac{1}{4}(\frac{1}{4} + \frac{3}{4}e^{-4/3\mu t})]^4[\frac{1}{4}(\frac{1}{4} - \frac{1}{4}e^{-4/3\mu t})]^2$. The first element, raised to the power of four, refers to the four sites that are unchanged, and the second element, raised to the power of two, refers to the two sites that differ between the taxa. Using this formula, we can then consider the likelihood that different branch lengths (values of μt) gave rise to the data in Figure 8.16. Looking at Figure 8.17, we observe that a branch length of $\mu t = 0.44$ has the highest likelihood (0.595×10^{-6}). This shows us that, under the JC model, 0.44 is the best estimate of the length of the branch in Figure 8.16.

Instead of keeping track of the likelihood, it is conventional to record the natural logarithm of the likelihood, the ***log-likelihood***. Using logarithms is helpful to avoid computer problems associated with handling very small numbers. A likelihood of 0.595×10^{-6} corresponds to a log-likelihood of -14.33 ($e^{-14.33} = 0.595 \times 10^{-6}$). The objective of maximum likelihood analysis is to find the tree with the highest likelihood. This corresponds to the least negative log-likelihood, which is -14.33 for the data in Figure 8.16.


Likelihood (× 1,000,000)
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0 1.1 1.2 1.3 1.4 1.5

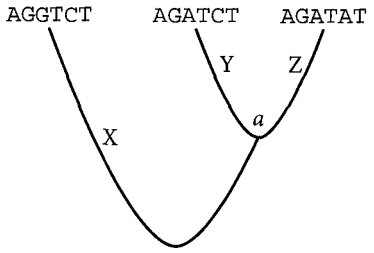


FIGURE 8.18 Three-taxon tree with a six base-pair sequence at each tip. The only items of uncertainty are the three branch lengths (X, Y, and Z) and the sequence at internal node *a*. We use the maximum likelihood criterion to estimate the value of the branch lengths, while summing over all possible sequences at node *a*.

the likelihood. Eventually, the program finds a set of values that cannot be significantly improved upon, and concludes the search.

The problem is even more challenging with larger numbers of taxa, because adding taxa not only adds branches whose lengths need to be considered but also adds ancestors whose character states are unknown and have to be inferred. For example, let us expand the previous example to add a third taxon (Figure 8.18). We now have three branches (X, Y, and Z) instead of just one, and we have created a node, *a*, whose sequence is unknown. The calculation of the likelihood looks at a branch and asks, What is the probability of the data that are observed at the two ends? But what if you have only observed data at one end of a branch?

Although we cannot observe the identity of the base at node *a* for any of the six characters, for each character we know that it was one of the four bases, A, C, G, or T. This means that we can determine the site likelihoods by summing over the four possible states at node *a*. We take the sum, rather than the product, because the observed data could have evolved because node *a* had state A *or* C *or* G *or* T. However, as before, we multiply the site likelihoods to obtain the overall likelihood. Under the Jukes-Cantor model, the maximum likelihood estimates of the lengths of the three branches in Figure 8.18 are X = 0.188, Y = 0, Z = 0.188. The log-likelihood of this tree is −15.92.

The computational challenges get even harder as we add additional taxa. With four taxa we have two internal nodes and, thus, 4^2 (16) possible sets of ancestral states that we need to sum over to determine the site's likelihood. With five taxa there are 4^3 (64 histories). In general, there are $4^{(n-2)}$ histories, where n is the number of taxa in the tree. The necessity of summing over all

these histories is one of the principal reasons that maximum likelihood analysis is so computationally intensive. Fortunately, advances in computation and some creative shortcuts have made these calculations practical even for very large data sets.

To recap: computer programs that perform phylogenetic analysis using maximum likelihood follow four steps. First, a particular tree and parameters (including branch lengths) are set and the likelihood of each site (character) is determined by summing over all possible histories for that site. Second, the likelihoods of each site in the matrix are multiplied (their logarithms are added) to obtain the overall likelihood of *that* tree with *those* parameters and branch lengths. Third, the program optimizes the branch lengths and other parameters by changing them iteratively and repeating the first two steps until the likelihood is maximized. This gives the maximum likelihood estimates of these parameters for the first tree. Fourth, a search through tree space is conducted to find the maximum likelihood tree, the tree on which the probability of the data is maximized. For every tree considered in the last stage in this process it is necessary to propose a set of branch lengths, calculate the likelihood, and then iteratively optimize the branch lengths. This should yield the tree that has the highest likelihood.

It is probably apparent that the computational challenge of maximum likelihood is great even for a simple model of molecular evolution. This is why there was such a lag between the 1970s, when the maximum likelihood method was first applied to phylogenetic inference (Felsenstein 1973), and the mid-1990s, when it became feasible to apply the method to real data sets. In the intervening years, computers became much faster and computer scientists and theoreticians found some effective shortcuts for doing likelihood calculations. As a result, it is now possible to do a full likelihood analysis of a data set like the carnivorans in less than an hour on a personal computer. The maximum likelihood estimate of the optimal tree for the carnivoran molecular data (using the HKY model of evolution) has a log-likelihood of approximately −5184. As shown in Figure 8.19, this tree is similar to the neighbor-joining tree but differs in the position of the bear and hyena lineages.

Currently available models of morphological evolution are rather simplistic. For example, they typically assume that all characters have the same rate of evolution and that the only character states a character can adopt are those that were observed in at least one of the taxa. Nonetheless, it is still possible to apply maximum likelihood to morphological data. Figure 8.20 shows the results of a maximum likelihood analysis of the carnivoran morphological data. This tree has an identical topology to one of the two maximum parsimony trees.

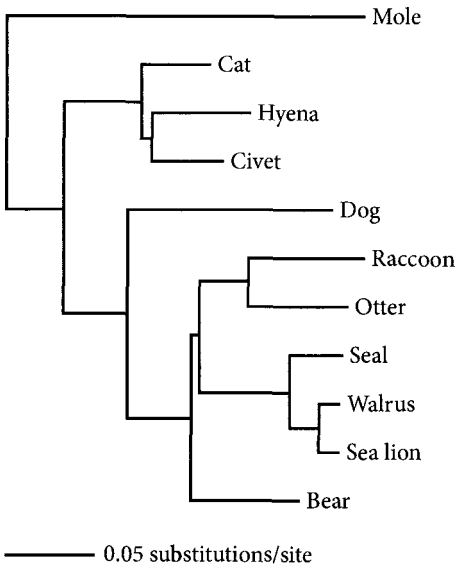


FIGURE 8.19 Maximum likelihood tree based on the carnivoran molecular data. A scale bar is provided to indicate branch length.

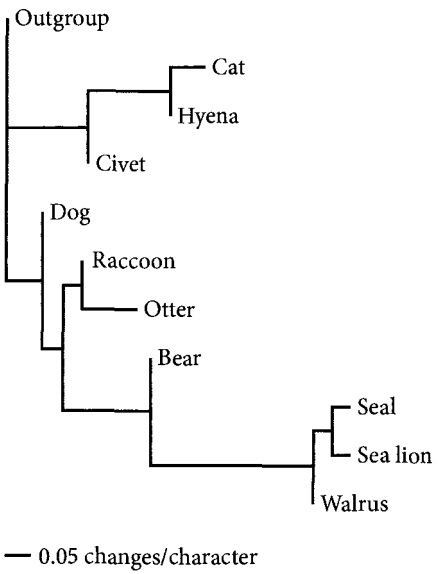


FIGURE 8.20 Maximum likelihood tree based on the carnivoran morphological data. A scale bar is provided to indicate branch length.

CHOOSING AN EVOLUTIONARY MODEL

As we have seen, the maximum likelihood criterion provides a way to estimate the value of any parameter in a model. Of particular note, base frequencies and the relative rates of different kinds of substitutions are estimated from the data at the same time as a tree is being inferred. This differentiates model-based methods from generalized parsimony (Chapter 7), which lacks any objective way to select the right costs to apply to different kinds of changes. Generalized parsimony asks the practitioner to assert whether there is a transition:transversion bias and, if so, how much of a cost should be used to reflect this fact. In contrast, under maximum likelihood, the data themselves are used to select the parameter values that maximize the probability of obtaining the observed data.

While maximum likelihood picks the parameters of the model (e.g., the base frequencies in F81, the transition:transversion bias in HKY), how do we pick the class of model to use? How do we decide whether to use JC, F81, HKY, or GTR? Likewise, how do we decide whether or not to allow for rate variation across sites and whether a molecular clock applies? It turns out that we can actually use the data to guide the choice of models. To get a feel for how this works, it is important to understand how the likelihood will change as you add more parameters to a model.

As discussed earlier, models with fewer parameters can be understood to be special cases of parameter-rich models (see Figure 8.10). The simpler model is said to be ***nested*** within the more complex one. When this is the case, the more complex (parameter-rich) model is *guaranteed* to have a likelihood that is equal to or greater than the simpler model. This makes sense. Each parameter added to a model makes the model more able to adapt to features of the data. In the same way that a metal chain with many smaller links can wrap itself more precisely around a pole than a chain with a few large links, a model with more free parameters can describe the data more precisely and will therefore yield a higher likelihood than a simpler model.

You might think that because more complex models always yield higher likelihoods, we would always use the most complex model we can define. However, there are costs to using an excessively complex model. A more complex model has more parameters and inevitably slows down the analysis. Also, in much the same way that a chain of a given length with more links is weaker than one with fewer links, more complex models come at the cost of reduced

statistical power. It may be impossible to say if tree 1 is significantly better than tree 2 when using an overly parameter-rich model, while a more appropriate model might allow us to conclude with confidence that tree 2 is false. When it comes to parameters, we *can* have too much of a good thing.

If there are costs to overly simple models (inability to account for the processes underlying the data) and to overly complex ones (losing the power to distinguish models), how do we choose which model to use? The principle is simple enough: pick a more complex model only when the gain in likelihood is more than would be expected if the simpler model were true. For example, if evolution followed the JC model, we would get a higher log-likelihood when we assumed the F81 model, but it should not be *much* higher than if we had assumed JC. If it were, we would have grounds to suspect that the underlying base frequencies really were different from 25%.

Statisticians have developed ways to predict how much higher the log-likelihood would be under a more complex model if the simpler were true. Different approaches use different formulae for calculating the expected likelihood gain due to adding extra parameters. One common approach (***hierarchical likelihood ratio tests***) can be used to compare nested models, where the more complex model has p extra parameters. Readers with a background in statistics may be interested to know that twice the log-likelihood difference between the two models is expected to fit a chi-square distribution with p degrees of freedom. For example, recalling that the F81 model has three more free parameters than the JC model (Figure 8.10), there is only a 5% probability that the F81 model will be more than 3.9 log-likelihood units higher than the JC model if the JC model is in fact true. This is because, under the χ^2 distribution, a value of ~7.8 (= 3.9 × 2) corresponds to a P-value of 0.05 for 3 degrees of freedom. Thus, if for a particular data set F81 yields a log-likelihood that is 4 or more units higher than JC, then we can be confident that the assumption of equal base frequencies is violated.

BAYESIAN INFERENCE

The newest approach to estimating phylogenies is Bayesian inference. Whereas likelihood judges a tree based on how probable it is that evolution would have produced the observed data, Bayesian inference judges trees based on their posterior probability, the probability that the tree is true, given the data, our

models of evolution, and our prior beliefs. To give you a feel for the Bayesian approach, let us return to the coin example introduced in the last section.

Recall that you were given a bag of coins, half of which are fair (50% heads) and half of which are biased (75% heads). You pull a coin out of the bag and want to know if it is biased or not. Before even tossing the coin, the ***prior probability*** of it being biased is 0.5 because we know that half the coins in the bag are biased. After tossing the coin, we can use the results to update this probability. Because this extra information comes after we have collected data, it is called the ***posterior probability***.

Recall that we were able to deduce previously that having a biased coin was more likely to generate the observed outcome than a fair coin, but we did not actually calculate the probability that the coin was biased. The principles for calculating such posterior probabilities were developed by the Reverend Thomas Bayes in the 18th century. He proved mathematically that the probability of a hypothesis, given some data, is equal to the probability of the data, given the hypothesis (the likelihood), times the prior probability of the hypothesis and divided by the probability of the data (summed over all hypotheses). Or more formally:

$$\text{Bayes' theorem:}\quad \underbrace{\Pr(H|D)}_{\text{Posterior probability}} = \frac{\underbrace{\Pr(D|H)}_{\text{Likelihood}} \times \underbrace{\Pr(H)}_{\text{Prior probability of the hypothesis}}}{\underbrace{\Pr(D)}_{\text{Prior probability of the data}}}$$

In this equation, Pr refers to "probability," D to "data," and H to "hypothesis." The vertical line is read as "given." For example, Pr(H|D) should be read as "the probability of the hypothesis, given the data."

We can apply this equation to obtain the posterior probability of the coin being biased. The probability of the data (10 heads), given the hypothesis that the coin is biased, Pr(D|H), is the likelihood. Using the same model as previously, this is 0.75^{10}, or 0.0563. The prior probability, Pr(H), that the coin is biased is 0.5. So the numerator in this example is 0.0281.

What is the prior probability of the data, the denominator in Bayes' equation? You might imagine that this probability is 1.0, seeing as the coin must be either fair or biased. But remember that we are trying to determine the probability of getting 10 heads in 10 tosses, as opposed to 9 heads and 1 tail, 8 heads and 2 tails, and so on. This probability will certainly be less than 1.0.

To calculate the prior probability of the data, Pr(D), we need to determine the probability of obtaining those data under each hypothesis and then we need to sum over all possible hypotheses (weighted by the hypotheses' prior probability). Here, there are only two possible hypotheses, namely, that the coin is biased or that the coin is fair. There is a 0.5 chance it is fair, and if it is fair the probability of getting 10 heads is 0.5^{10}. There is a 0.5 chance that it is biased, and if it is biased the probability of getting 10 heads is 0.75^{10}. Summing these together, the probability of the data is $0.5 \times (0.5^{10} + 0.75^{10}) = 0.0286$. This means there was a 2.9% chance that you would have grabbed a coin at random from the bag and then obtained 10 heads in 10 tosses.

Combining these numbers, the posterior probability that the coin is biased after observing 10 heads in a row is 0.0281/0.0286, or about 0.98. This means that, given the priors and model, there is a 98% chance that the coin is biased and only a 2% chance that it is fair. By taking into account the observations of 10 consecutive heads, the probability that you drew a biased coin has jumped from 0.5 to 0.98, while the probability that you drew a fair coin has dropped from 0.5 to 0.02.

What is interesting and special about the Bayesian approach is that the starting information matters. Suppose, for example, that you drew the coin from a sack that had only 1% biased coins. In that case the posterior probability that it is biased after getting 10 heads in a row is only 0.37. While the data has moved you from a posterior probability of 0.01 to 0.37, the posterior is still less than 0.5. This means that even after observing 10 heads you would still be wise to bet against it being a biased coin!

In this coin example, the alternative hypotheses each indicated an exact probability of heads, and hence a simple calculation of the likelihood of the data. What if you knew that the sack contained two kinds of coins at equal frequency: "fair" coins whose probability of yielding heads is between 0.4 and 0.6 and biased coins whose probability of heads is somewhere between 0.7 and 0.9? What would you do? By extrapolation from the discussion of maximum likelihood, you might argue for selecting whatever value within these ranges maximizes the likelihood. Thus, if you observed 10 heads you would select a value of 0.6 for the fair coin hypothesis (likelihood = 0.6^{10}) and 0.9 for the biased coin hypothesis (likelihood = 0.9^{10}). However, this is *not* what the Bayesian approach calls for. Instead, a good Bayesian would integrate over all possible values of the parameter, weighted by the prior probability of each being the true value of the parameter. This is one reason why Bayesian methods can

yield different conclusions than maximum likelihood, even in cases where all hypotheses have the same prior probability.

Now let's apply these principles to phylogenetics. The data correspond to a character state matrix and the hypotheses correspond to the alternative possible tree topologies. Thus Bayes' theorem takes the following form:

$$\Pr(\text{Tree}|\text{Data}) = \frac{\Pr(\text{Data}|\text{Tree}) \times \Pr(\text{Tree})}{\Pr(\text{Data})}$$

The prior probability of a particular tree topology, Pr(Tree), is the probability (before looking at your data) that among all possible trees it is the true tree. For example, if we believed that all tree topologies were equally likely *a priori*, we could apply a ***flat prior***, where the prior probability of each tree equals one divided by the number of distinct tree topologies (see Table 7.8).

Calculating the probability of the data given the tree, Pr(Data|Tree), entails determining the likelihood of the tree. This is done as described earlier in this chapter, except that instead of selecting the values of the free parameters (e.g., base frequencies, branch lengths) that maximize the likelihood of the data, we integrate over the prior probability for all parameters. These are often difficult calculations, but they are possible in many cases.

The big challenge with Bayesian phylogenetics is calculating the prior probability of the data, Pr(Data). The problem is that Pr(Data) involves a summation over all trees, but, as shown in Table 7.8, there are lots of possible tree topologies. Whereas a single tree in isolation can be assigned a parsimony score or likelihood, a Bayesian posterior probability cannot be assigned to a single tree without taking account of all other possible trees. As a result, Bayesian phylogenetics would be impossible without the invention of a clever method called ***Markov chain Monte Carlo*** (MCMC) analysis.

The MCMC method exploits the fact that while we cannot easily calculate the actual posterior, we can calculate the relative posteriors of different trees. To help you visualize this, imagine that you want to survey the altitude of every point in a landscape and you know that the true altitudes, summed over all points, is 100,000 m (this is analogous to knowing that the sum of posterior probabilities of all trees must sum to 1.0). You have a defective altimeter—it accurately measures altitudes except that each measurement is multiplied by some unknown constant. If you get a reading of 400 m at one point, you have no idea whether the real height is 40 m or 4000 m. This device is not totally use-

less, however. If you measured a second point at 360 m, you would know that it was 10% lower than the first point (the ratio of the two heights is 10:9). Thus, if you surveyed the entire landscape with the faulty altimeter, you could recalibrate all the measurements so that they summed to 100,000 m.

When applying MCMC to phylogenetic inference, the landscape is a highly multidimensional parameter space. In addition to containing all possible tree topologies, this space includes the range of possible branch lengths and all the free parameters of the model of evolution. This is a much more complicated space than the tree space introduced in Chapter 7, but the principle is identical. The process goes on within a computer, but we can imagine walking through this space ourselves. Figure 8.21 provides a visual representation of an MCMC run.

We start at some point in this parameter space and calculate the likelihood, Pr(D|H), the probability that evolution would have yielded exactly these data. Given our model of evolution and the parameters specified by our location in parameter space, we can calculate relatively easily the likelihood associated with this point in the landscape.

The next step is to randomly propose a new parameter combination and calculate its posterior probability. We also need to calculate something called the

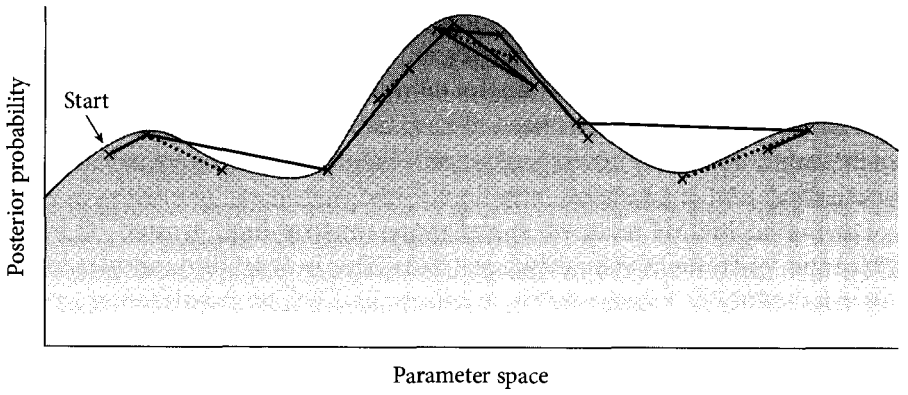


FIGURE 8.21 A visual representation of the Markov chain Monte Carlo method. Each × marks a parameter value that was proposed during the chain. Those proposals that were accepted are marked with a solid line, whereas those proposals that were rejected are marked with a dotted line. All "uphill" proposals are accepted. Some downhill proposals are rejected and some are accepted.

proposal ratio, but we will not describe this ratio (see Further Reading). The rule we will follow is that if the new point in parameter space is uphill, meaning it has a higher posterior probability, we will move to that new point. This higher posterior probability could arise because the new parameter combination yields a higher likelihood and/or because it has a higher prior probability. Either way, analogous to heuristic searching for optimal trees, we always accept a proposal that has a higher posterior probability (Chapter 7).

The difference between an MCMC run and a heuristic search algorithm is that when a proposal takes us downhill (i.e., when the product of the likelihood and prior is lower than the current state), we might nonetheless accept the proposal. We decide randomly (hence the Monte Carlo reference) whether to accept the "downhill" tree/parameter. The probability of accepting such a proposal is based on the ratio of the two posteriors. If there were a 10% difference between the posteriors, meaning the ratio is 10:9, there would be a 9/10 chance of accepting the proposal and a 1/10 chance of rejecting the proposal and thereby staying at the previous parameter value. The process of proposing a new parameter value and deciding whether to accept it is considered one ***MCMC generation***.

Starting from whichever parameter combination we chose in the last generation, we then initiate a new generation. A new proposal is made and the same rule is followed to decide whether the proposal should be accepted or rejected. The MCMC run continues for millions of generations, each consisting of a proposal that is either accepted or rejected. During this process, the computer keeps a list of the trees and parameter combinations that were visited during the chain (actually, we do not need to keep data for every generation but only a subsample, e.g., every 100th generation). Figure 8.21 may give you a feel for what a small part of an MCMC run "looks" like.

During an MCMC run, we spend more time on high ground (at better trees) and less time in valleys (worse trees). This is because it takes us longer to walk downhill (because we often reject downhill proposals) than to walk uphill (because we always accept uphill steps). It turns out that the frequency with which we find ourselves in a region of parameter space will eventually be proportional to that region's posterior probability. It can be mathematically proven that if we wander around the parameter space long enough, the proportion of trees in our list having a specific topology will be proportional to that topology's posterior probability. Similarly, for all parameters of our model (e.g., branch lengths, transition:transversion bias, rate heterogeneity across sites

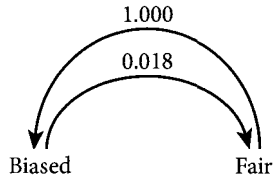


FIGURE 8.22 Representation of a Bayesian MCMC analysis in the coin example. The parameter space has two points, "Fair" and "Biased," corresponding to the two hypotheses. When the chain is on "Fair," the proposal of a switch to "Biased" will always be accepted. When the chain is on "Biased," there is a probability of only 0.018 that the proposal of a switch to "Fair" will be accepted.

in the sequence, etc.) the list of values visited during the MCMC run should approximate their posterior probability distribution.

To illustrate the idea, let us revisit the coin example and see if an MCMC approach will yield the same posterior probability as we obtained by calculation. Imagine two spots on the floor, one representing the case of a fair coin and one representing the case of a biased coin (with a 0.75 probability of yielding heads). We can use MCMC to calculate the posterior probability of the two hypotheses given that they each have a prior probability of 0.5 (Figure 8.22). We will start by stepping on one of the two spots. Then we will "propose" the other spot. The likelihood is higher under the biased coin hypothesis (−0.056) than under the fair coin hypothesis (−0.001). As a result, whenever we are on the spot representing the fair coin hypothesis, the proposal to switch to the other spot (representing a biased coin) will be accepted. In contrast, if we are on the biased coin spot, the probability of accepting the fair coin spot is defined by the ratio of the two likelihoods, meaning that the probability of accepting this proposal is 0.001/0.056 = 0.018 (Figure 8.22).

Imagine using MCMC to guide a decision as to whether to jump between these two spots. When you are on the biased spot, you will reject the proposal to move to the fair spot ~98% of the time and accept it ~2% of the time. If you accept the proposal, you will spend exactly one generation on the "fair" spot before jumping back to the "biased" spot. It should be obvious that after many generations, you will have spent a total of 98% of the time on the biased spot and 2% on the fair spot. As you will see, these values correspond to the posterior probabilities that we calculated earlier.

Bayesian phylogenetics using MCMC involves several complications, some of which are worth mentioning because they influence the performance of the method and the reliability of the output. First, it is necessary to select a model of trait evolution. Analogous to maximum likelihood analysis, the model may be selected prior to the MCMC analysis using hierarchical likelihood ratio tests and similar methods. Alternatively, it is possible to place prior probabilities on different models and allow the MCMC to jump between models, thereby integrating over uncertainty in model choice.

The second complication with MCMC is that the chain needs to reach ***stationarity*** before it provides useful information on the posterior probability. In this context, stationarity means that the likelihood is bouncing around but not showing any consistent upward trend. If we start at a very poor set of parameter values, it may take a long time to find our way to high ground, and we would not want the initial period spent at parameter values with low posterior probabilities to distort our conclusions. The general procedure is to identify the period before stationarity has been reached, the so-called burn-in, and delete these entries from the list. This is often done by looking at a plot of likelihood as a function of generation and deleting all generations before the point at which the curve levels out. Figure 8.23 shows an example of such a plot.

A third issue is to ensure that, during the period of stationarity, the run is exploring all of parameter space—that it is ***mixing***. If chains are mixing prop-

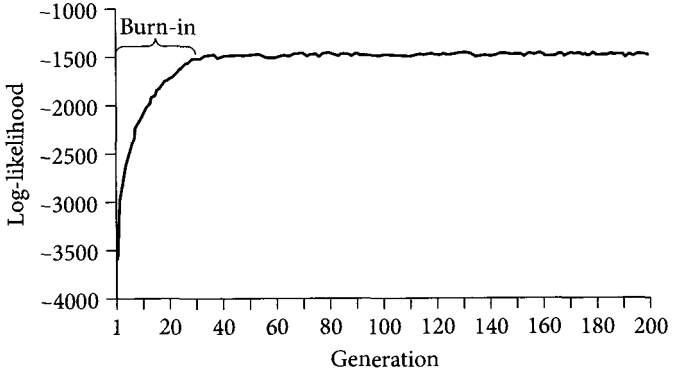


FIGURE 8.23 Change in likelihood during a Bayesian MCMC. Samples collected during the first part of the chain, when there is a steady increase in the likelihood, the burn-in, should be discarded before the posterior sample is summarized.

erly, then different runs will converge on a similar posterior distribution. This can be checked by initiating multiple runs from different random starting points and seeing if each run supports the same phylogenetic conclusions. If we find evidence of poor mixing, the MCMC procedure can be adjusted to improve its performance. Most adjustments include modifying the way that new points in parameter space are proposed. For example, sometimes each run will entail several coupled chains, where only one (the "cold" chain) is following the MCMC rules. In such a strategy, the other ("heated") chains follow modified rules of decision making and serve as a source of points in parameter space that can be proposed to the cold chain. More information can be found in the recommended Further Reading.

Once we are convinced that we have adequate mixing and thus a reasonable sample of the posterior probability distribution, we can query it to learn about particular parameters, for example, tree topology. By counting how many times a particular tree topology is sampled in the post–burn-in trees, we can obtain an estimate of that tree's posterior probability. For example, if our distribution contained 8000 trees of which 7200 have the same topology, that tree's posterior probability would be estimated to be $7200/8000 = 0.90$.

While Bayesian phylogenetics is complex and computationally demanding, the method has become relatively easy and quick to implement thanks to the development of user-friendly programs, such as MrBayes. This program can be run on various kinds of computers and is very flexible. The book *Phylogenetic Trees Made Easy* offers an introduction to the hands-on aspects of using this program for phylogenetic research (Hall 2011).

Analysis of the carnivoran molecular data set using MrBayes yields a posterior distribution containing nine distinct topologies. The tree with the highest posterior probability (0.76) is shown in Figure 8.24. The posterior probability of 0.76 means that, given our prior assumptions about the probability of different trees and models of molecular evolution (which we are glossing over), there is a 76% chance that this tree is correct and a 24% chance that it is wrong in some way.

The morphological data are less supportive of any one topology. The tree that has the highest posterior probability is shown in Figure 8.25. It has a posterior probability of only 7.9%. This low number is probably due to the fact that the morphological data set is small relative to the number of possible trees. With little information, tree space becomes flatter and the MCMC wanders more, spending less time visiting any particular tree.

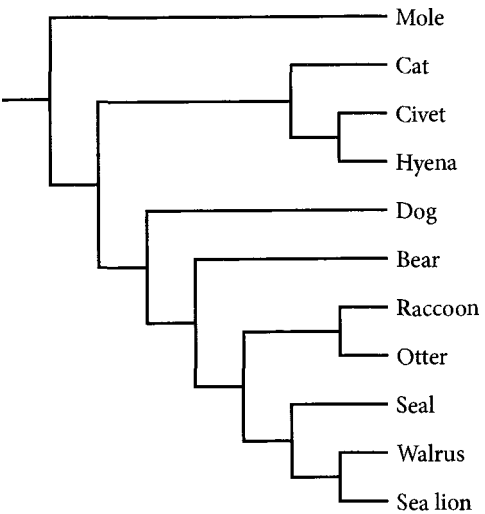


FIGURE 8.24 Tree topology with the highest posterior probability from a Bayesian MCMC analysis of the carnivoran molecular data. This analysis used the GTR+Γ model of evolution.

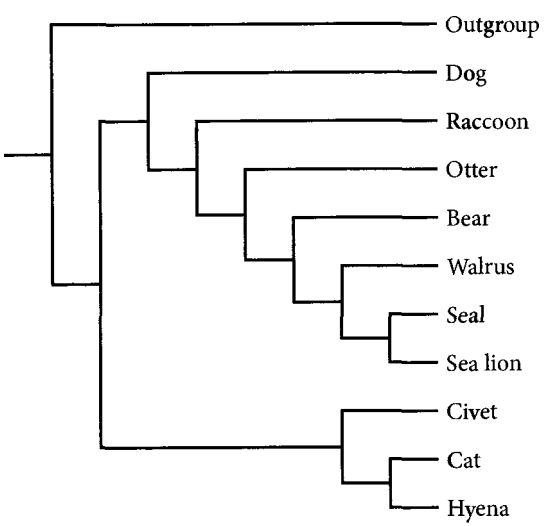


FIGURE 8.25 Tree topology with the highest posterior probability from a Bayesian MCMC analysis of the carnivoran morphological data.

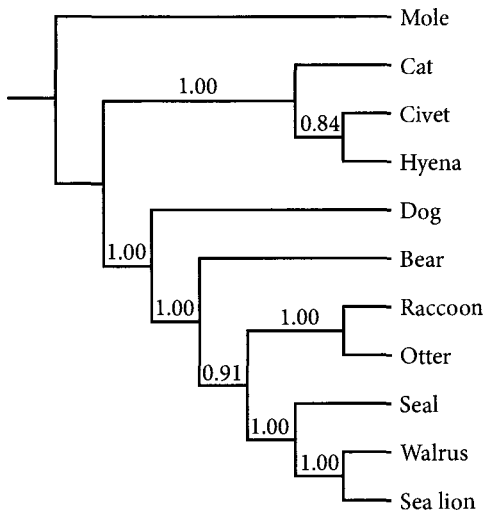


FIGURE 8.26 Bayesian majority-rule consensus tree for the carnivoran molecular data. Numbers on branches are clade posterior probabilities: the fraction of trees in the posterior sample that contain the clade in question. This analysis used the GTR+Γ model of evolution.

Generally, in phylogenetic analysis, we are not specifically interested in the posterior probability of a particular tree topology. Rather, we care about the posterior probability of individual clades. We can calculate the posterior probability of a particular clade by seeing how often it appears during the MCMC analysis. For example, if a clade is present in 7600 of 8000 trees, then its posterior probability (sometimes called its ***clade credibility***) is 0.95. Typically, a posterior distribution is summarized by drawing a tree composed of clades whose posterior probability is greater than 0.5. This is the ***Bayesian majority-rule consensus tree***, often loosely called a ***Bayesian tree***.

The Bayesian majority-rule consensus tree for the carnivoran molecular data obtained with MrBayes is shown in Figure 8.26. It has the same topology as the single tree with the highest posterior probability. The numbers on each branch are clade credibility estimates. This analysis suggests that all but two clades have a greater than 95% probability of being true (given the model and priors).

The Bayesian consensus tree for the morphological data is shown in Figure 8.27. In contrast to the molecular tree, only three clades have credibility scores greater than or equal to 0.95. Thus, once again we see that the carnivoran

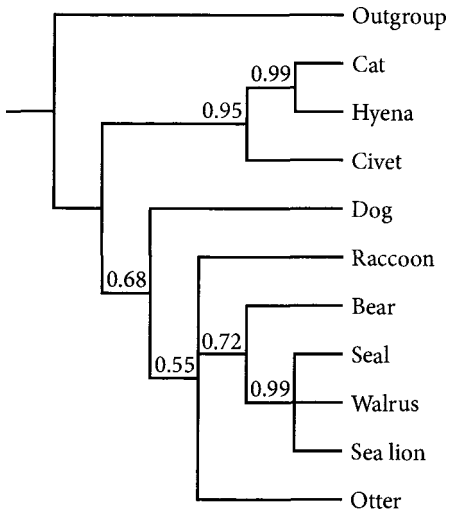


FIGURE 8.27 Bayesian majority-rule consensus tree for the carnivoran morphological data. Numbers on branches are clade posterior probabilities. Branches with posterior probabilities < 0.50 have been collapsed into polytomies.

molecular data set contains more phylogenetic signal than does the morphological data set.

Bayesian phylogenetic analysis has a lot of desirable features, not least of which is that the endpoint is something that scientists usually want, namely, an estimate of the probability that a tree is true (given the data, model, and priors). Nonetheless, some practitioners are uncomfortable with the approach, primarily because of doubts over the very structure of Bayesian statistics. In particular, some phylogeneticists object to having to specify an exact prior probability distribution for all parameters. This, they feel, makes the approach subjective—if you and I have a different set of prior beliefs, the posterior probabilities we obtain will be different. However, one can argue that all of science involves updating probabilities of hypotheses by combining prior knowledge with new data. When we conduct any experiment, we apply some prior knowledge about the system to guide the collection and interpretation of new data. Bayesian inference simply provides a formal way to combine prior knowledge with new information. Given the computational tools and the philosophical appeal of Bayesian inference, it is perhaps not surprising that Bayesian inference is becoming the most widely used method for phylogenetic analysis.

FURTHER READING

General resources: Swofford et al. 1996; Felsenstein 2004
Models of evolution: Swofford et al. 1996; Lewis 1998, 2001
Distance methods: Fitch and Margoliash 1967; Felsenstein 1984; Saitou and Nei 1987; Rzhetsky and Nei 1992
Maximum likelihood: Felsenstein 1981a; Huelsenbeck and Crandall 1997; Lewis 1998
Model selection: Goldman 1993; Posada and Crandall 2001
Bayesian phylogenetics: Larget and Simon 1999; Mau et al. 1999; Huelsenbeck et al. 2002; Holder and Lewis 2003; Huelsenbeck et al. 2004; Lewis et al. 2005

CHAPTER 8 QUIZ

1. In the substitution probability matrix for JC (Figure 8.5), the top right entry is $\frac{1}{4} - \frac{1}{4}e^{-4/3\mu t}$. What does this mean? ($\mu$ is the substitution rate; t is time)
   a. The rate of going from A to T is $\frac{1}{4} - \frac{1}{4}e^{-4/3\mu}$ changes per unit time, t.
   b. The proportion of changes that occur during a time window that are from A to T is $\frac{1}{4} - \frac{1}{4}e^{-4/3\mu t}$.
   c. The probability of starting at A and ending at T at each site in the sequence is $\frac{1}{4} - \frac{1}{4}e^{-4/3\mu t}$.
   d. Answers a and b are correct.
   e. Answers a, b, and c are correct.

Questions 2–3. Assume that a gene evolves according to the JC model with a substitution rate of 5.2×10^{-10} substitutions per site per year. (*Hint*: Refer to Figure 8.5 and use a calculator or spreadsheet.)

2. A particular site is an A now. What is the probability that this base will be a G after 25 million years?
   a. 0.9871 b. 0.0617 c. 0.0129 d. 0.0043 e. 0.0011

3. A particular site is an A in species X and a G in species Y. What is the site likelihood under the assumption that their last common ancestor lived 12.5 million years ago?
   a. 0.9871 b. 0.0617 c. 0.0129 d. 0.0043 e. 0.0011

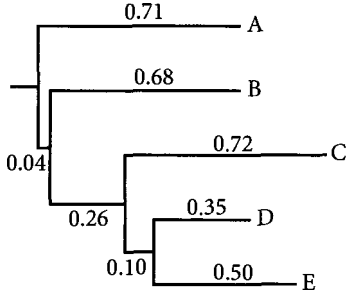
4. Given the following data matrix, what is the pairwise distance between A and C?

	1	2	3	4	5	6	7	8	9	10
A	1	0	0	0	1	1	1	0	1	0
B	0	1	0	1	1	1	1	0	0	0
C	0	0	1	1	0	1	0	0	1	1

a. 0.1 b. 0.4 c. 0.6 d. 2 e. 4

5. What is the best description of the neighbor-joining method?
   a. A tree is estimated based on how close taxa are in a data matrix.
   b. A tree search is conducted to find the tree that minimizes the difference between observed and predicted distances.
   c. The shortest tree is found under the assumption that there is a shortest tree.
   d. A protocol for adding taxa to a backbone tree built by parsimony or maximum likelihood.
   e. A fast algorithm that will find a tree based on pairwise distances.

6. Given the tree, what is the evolutionary distance between taxa B and D?



a. 0.33 b. 1.03 c. 1.30 d. 1.39 e. 1.43

7. What is the minimum evolution method?
   a. Another name for parsimony.
   b. A variant of neighbor-joining that minimizes the tree length.
   c. A method for picking among tree topologies (when the trees' branch lengths have been optimized by the least-squares method).
   d. A method for determining the branch lengths that would be most likely to generate the observed data given a tree topology.
   e. A method for subtracting the observed and expected distances to best explain character evolution.

8. Which of the following is the best equivalent of likelihood?
   a. The probability that an event will occur.
   b. The probability of observing a set of data given a hypothesis.
   c. The probability that a hypothesis is true given a set of data.
   d. The ratio of the probabilities of two alternative hypotheses.
   e. The probability that the model fits the data.

9. We have run a likelihood search for a molecular data set, and we find that a particular tree with branch lengths has a log-likelihood score of −2467. What does this number mean?
   a. The probability of this data matrix evolving along this tree is e^{-2467}.
   b. The probability of the individual characters in the data set sums to 10^{-2467}.
   c. The probability that the tree is true is e^{-2467}.
   d. The tree is 2467 times more likely to generate the data than a random tree.
   e. This tree requires 2467 fewer steps than the longest tree we can find.

10. Which of the following is true of posterior probabilities? (select all that apply)
    a. They can be calculated using Bayes' theorem.
    b. They do not depend on the choice of a model.
    c. They incorporate prior information about the probability of alternative hypotheses.
    d. They can change as additional data are collected.
    e. Hypotheses with smaller posterior probabilities are favored over those with larger values.

11. In a particular population, we know that 60% of the frogs are female and the rest are males. All male frogs sing, but only 10% of the female frogs sing. Walking through the population, you hear a frog singing. What is the posterior probability that the singing frog is a male? [*Hint*: The probability of any frog singing is the probability of a singing female (0.6 × 0.1) plus the probability of a singing male (0.4 × 1.0) .]
    a. 0.23 b. 0.24 c. 0.40 d. 0.46 e. 0.87

12. During an MCMC analysis, what determines when a chain moves from the current tree to a new tree? (select all that apply)
    a. The chain will always accept a proposal that yields a higher posterior probability than the current state.
    b. The chain will always accept a proposal that is consistent with the prior.
    c. The chain will accept a proposal with a probability that is fixed at the start of the search.
    d. The chain will reject a proposal that yields a lower log-likelihood than the current state.
    e. The probability of accepting a proposal is dependent on the ratio of the posterior probabilities of the current and proposed states.

13. An MCMC run currently has a state whose likelihood is 10^{-22}. A new state is proposed whose likelihood is 10^{-23}. Assuming that the ratio of the priors is 1 (i.e., that the probability entirely determines the acceptance ratio), what is the likelihood that the new state will be accepted?
    a. 10^{-23} b. 0.1 c. 0.091 d. 0.49 e. 1.0

14. Below are 10 post–burn-in trees in parenthetical notation from a Bayesian MCMC search on a data set of five carnivores. Looking at just these trees, what is the posterior probability of the seal + walrus clade?

```
TREE 1 = (((otter,(seal,walrus)),bear),creodont);
TREE 2 = ((((seal,otter),walrus),bear),creodont);
TREE 3 = ((otter,(bear,(walrus,seal))),creodont);
TREE 4 = ((((bear,seal),otter),walrus),creodont);
TREE 5 = ((((seal,otter),walrus),bear),creodont);
TREE 6 = (((bear,(seal,walrus)),otter),creodont);
TREE 7 = ((((walrus,otter),seal),bear),creodont);
TREE 8 = ((otter,((seal,walrus),bear)),creodont);
TREE 9 = ((((walrus,seal),otter),bear),creodont);
TREE 10 = ((((otter,seal),walrus),bear),creodont);
```

    a. 0.3 b. 0.4 c. 0.5 d. 0.6 e. 0.7

15. Briefly describe what it means to obtain the likelihood of a tree given a data set.

16. We have compared the likelihoods for a particular data set and a single tree for four models of evolution in order to choose the best model for the final analysis. The log-likelihoods are given below:

    JC model: −1872.44
    F81 model: −1855.35
    HKY model: −1859.27
    GTR model: −1856.32

    One of these likelihoods was calculated incorrectly. Which? How do you know?

17. What are the main strengths and weaknesses of the neighbor-joining method?

18. Assume two trees have equal prior probability and that the likelihood of the data under tree 1 is twice the likelihood of the data under tree 2. Given these facts, can you calculate the *relative* posterior probabilities of trees 1 and 2 given these data? If the answer is yes, what is the relative posterior probability? Can you calculate the *absolute* posterior probability of tree 1? If the answer is yes, what is its posterior probability? If the answer to either question is no, why can't you?

19. Using tree space terminology, explain why trees visited more often in an MCMC search have higher posterior probability.

20. Consider the processes of phylogenetic inference with maximum likelihood and Bayesian analyses. Why might some practitioners prefer one method over the other?

CHAPTER NINE

Statistical Tests of Phylogenetic Hypotheses

Chapters 7 and 8 introduced the major methods by which phylogenetic trees are inferred, namely, maximum parsimony, neighbor-joining (the most widely used distance method), maximum likelihood, and Bayesian inference. With these methods, it is possible to obtain a good estimate of evolutionary relationships. However, regardless of which method is used, the optimal tree cannot be assumed to *be* the true tree. Even if all the assumptions of a method are met, there is an element of chance in trait evolution, which means that the data will sometimes suggest a tree that deviates to some degree from the true tree.

Statistical inference is now understood to be an intrinsic part of phylogeny reconstruction. This is why the terms "phylogenetic reconstruction" and "tree building" are somewhat misleading. They imply that trees are built according to some predefined procedure, at which point they are solid and fixed entities, like buildings. Rather, trees obtained by phylogeneticists are subject to error like all other scientific hypotheses and are more appropriately referred to as "estimates" of the true history. The field of phylogenetics should not be seen as an attempt to *build* trees, but rather to examine alternative trees and then quantify the extent to which data support or reject different phylogenetic conclusions.

In this chapter we explore how to establish confidence in various phylogenetic inferences. We start by considering the problem of determining whether there is any meaningful signal in a data set. Next, we consider several approaches that may be used to assess how confident we should be in particular clades or topological conclusions. The methods discussed include decay indices and bootstrap percentages, which may be used to annotate a tree diagram so as to communicate the relative support for different clades, and parametric bootstrap analysis, which may be used to formally test competing hypotheses. The

last few sections of the chapter relate to situations in which multiple data sets from the same taxa have been collected, and we wish to evaluate whether they have all tracked the same tree topology. This is significant because disagreement between genes can shed light on the evolutionary processes that have shaped the distribution of genetic variation among species.

DETECTING NONRANDOMNESS IN A DATA SET

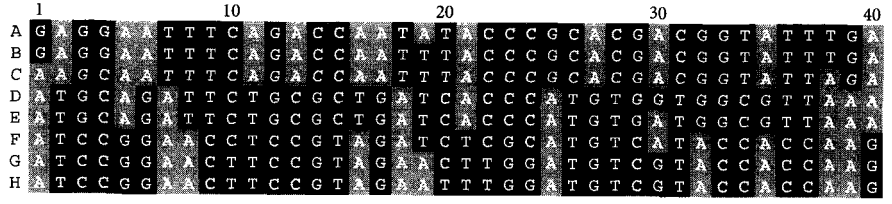
The starting point for phylogenetic analysis is usually the assumption that our data derive from a treelike descent process. However, a lack of treelike structure is the expected result if the lineages all diverged from their most recent common ancestor simultaneously at a hard polytomy, forming what is called a "star phylogeny." Even if the true tree is not a star, phylogenetically randomized data could result if, for example, the characters evolve so rapidly that the historical signal has been overwritten by subsequent changes.

Phylogenetic methods will usually yield a tree regardless of whether the data have any phylogenetic signal. Even if the data do not derive from evolution along a tree, some trees will, by chance, be more highly supported (by whatever criterion) than others. It is prudent, therefore, to develop ways to determine if a data set has a nonrandom, treelike structure.

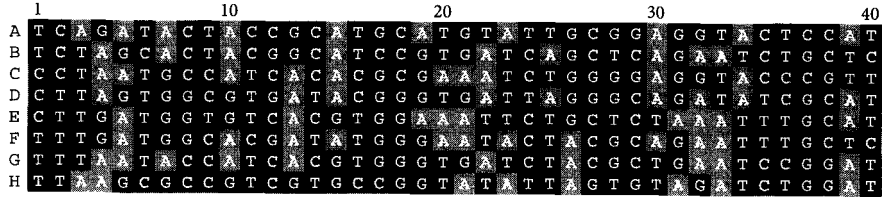
How do data derived from evolution along the branches of a tree differ from what we would expect if there were no tree signal? As discussed in Chapter 5, independent characters evolving slowly along a tree tend to define nested, that is, nonoverlapping, taxa (clades). Characters evolving along a tree will only conflict with one another, supporting overlapping taxa, when they experience homoplasy. So unless homoplasy has been rampant, trees should yield characters that conflict with one another less than do random data. Thus, we can detect treelike structure by looking at the extent to which characters within a matrix contradict each other.

Figure 9.1 illustrates this concept. The top data matrix shows the first 40 parsimony-informative characters that arose during simulated evolution of a DNA sequence up the tree shown. As you can see by eye, there is considerable structure in these data. For contrast, a randomized set of 40 characters is shown below. The latter clearly has much more inter-character conflict. But how can we quantify that conflict?

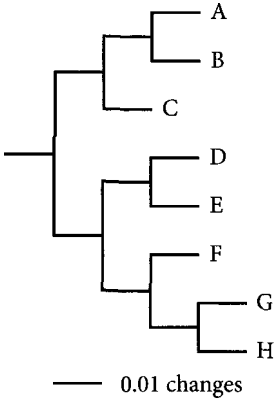
The two data sets contain an identical number of characters and character states, yet measures of the quality (length, likelihood) of the optimal tree are



a



b



c

FIGURE 9.1 Data derived from evolution along a tree are expected to look quite different from random data. Simulated evolution along the tree shown in (c) yielded the data set in (a). This contrasts with (b), which contains the same number of characters and the same distribution of character states as data set (a), but lacks any phylogenetic structure. Data set (a) shows much more agreement among characters than data set (b).

different. For example, the shortest tree that we can find for the tree-derived data, 43, is much shorter than the shortest tree for the random data, 89. Likewise, the log-likelihood of the optimal tree for the tree-derived data set (using the HKY model; see Chapter 8) is −192.25, much higher than the random data set, which is −347.66.

These differences are readily understood if we remember that characters evolving on a tree will tend to define hierarchically nested groups (recall Figure 5.11). When characters consistently support the same nested sets of clades, a tree that contains those clades will tend to be short and have a high likelihood. In contrast, when characters are random, no single tree provides a simple explanation for all the variation. This raises the question, How can we assess whether the optimal tree has a lower parsimony score or higher likelihood score than is expected for random data?

Several methods are available, of which we will introduce just one, the ***Permutation Tail Probability (PTP) test.*** This test was the basis for Table 2.2, which was presented to illustrate the idea that the presence of hierarchical structure in data provides concrete evidence for descent from common ancestry (as contrasted with separate ancestry). While this test can be conducted using any method that assigns a score to an individual tree (i.e., parsimony, minimum evolution, and likelihood, but not neighbor-joining), we will illustrate it using equally weighted parsimony because this is the quickest of the methods and is sufficient to determine whether data contain phylogenetic structure.

The PTP test compares the score of the shortest tree to the scores of the most parsimonious trees found with random data sets. In order to simulate phylogenetically random data, the original data matrix is ***permuted.*** This means that the states within a character are randomly reassigned to taxa, and this is done separately for all characters. Because character states within each character are shuffled independently of the other characters, any hierarchical structure that was present is obliterated. Table 9.1 shows two permutations of a five-taxon, five-character data set. Notice that character 1 always includes two A's and three C's, but the assignment of these to taxa varies.

In a PTP test, the data are permuted a large number of times (e.g., 1000) and the permuted data sets are subjected to phylogenetic analysis. We do not pay attention to the topology of the inferred trees but just record their length. When the length of the shortest tree for the original data is shorter than all or nearly all of the optimal trees obtained from the permuted data sets, the data can be said to have more phylogenetic structure than is expected for random data.

TABLE 9.1 Two random permutations of an original data set

Original data						Permutation 1						Permutation 2					
	1	2	3	4	5		1	2	3	4	5		1	2	3	4	5
A	A	G	T	C	T	A	C	A	C	G	T	A	A	G	T	A	A
B	A	G	T	C	T	B	A	G	C	C	A	B	C	A	C	C	T
C	C	G	T	C	A	C	C	A	T	A	C	C	C	G	T	C	C
D	C	A	C	A	A	D	A	G	T	C	A	D	C	A	C	C	A
E	C	A	C	G	C	E	C	G	T	C	T	E	A	G	T	G	T

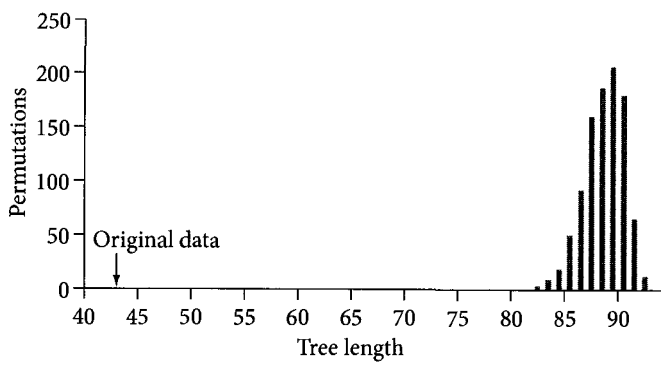


FIGURE 9.2 **PTP test using the treelike data from Figure 9.1.** The trees found with permuted data are consistently longer than the optimal tree with the original data. This shows that the original data have significant phylogenetic signal.

As an aside, you may notice that in Table 2.2 we followed the original author in using the consistency index (minimum tree length ÷ observed tree length; see Chapter 4) as a measure of hierarchical structure, which is to say, phylogenetic signal. Using tree length instead of the consistency index yields the same conclusions, however, because the consistency index is perfectly (albeit inversely) correlated with tree length.

To illustrate PTP analysis, the 40 informative characters in the upper part of Figure 9.1 were permuted 1000 times and, for each, the length of the most parsimonious tree was determined. As you can see, in Figure 9.2, none of the permuted (randomized) data sets yielded a tree as short as the original data set,

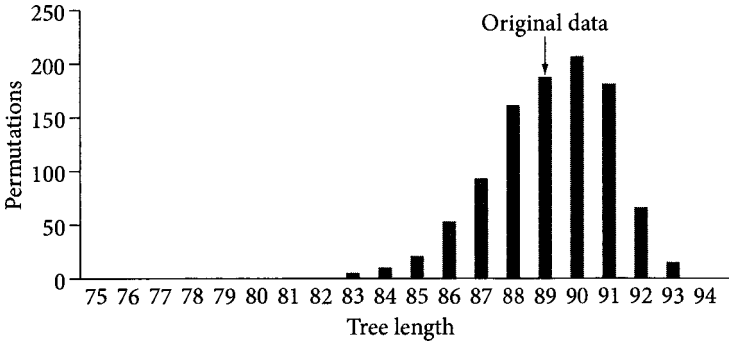


FIGURE 9.3 PTP test using the randomized data from Figure 9.1. The trees found with permuted data are not consistently longer than the optimal tree with the original data (from Fig. 9.1b). This shows that the original data do not have significant phylogenetic signal.

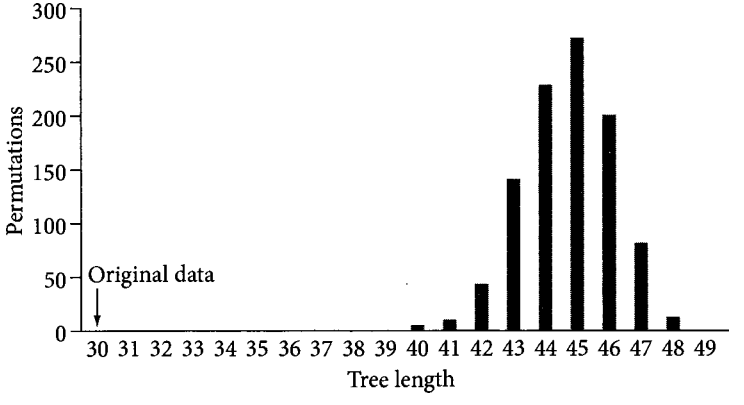


FIGURE 9.4 PTP test for the carnivoran morphological data. The trees found with permuted data are consistently longer than the optimal tree with the original data. This shows that the carnivoran morphological data have significant phylogenetic signal.

whose length was 43. This result shows that the original data have much more structure than we expect from random data.

In contrast, a PTP test for the randomized data in the lower panel of Figure 9.1 yields a different pattern (Figure 9.3). In this case the original data set corresponds to an optimal tree of length 89, which is similar in length to that of trees obtained from the permuted data sets. This confirms that these data did not derive from evolution along the branches of a phylogenetic tree.

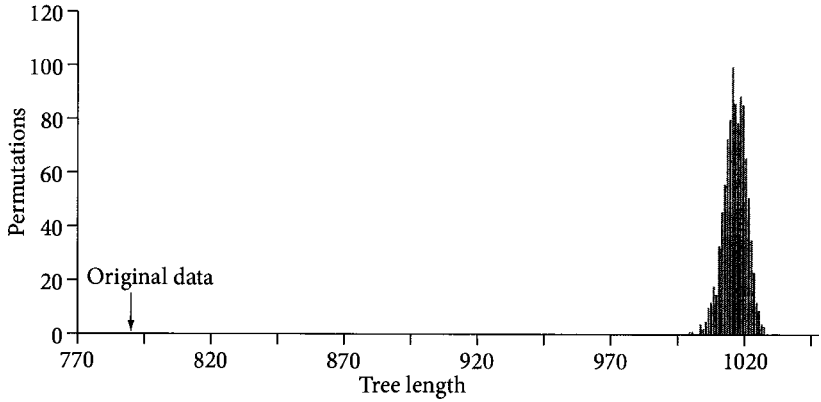


FIGURE 9.5 PTP test for the carnivoran molecular data. The trees found with permuted data are consistently longer than the optimal tree with the original data. This shows that the carnivoran molecular data have significant phylogenetic signal.

Figure 9.4 applies the PTP test to the carnivoran morphological data (see Chapter 7). The shortest trees for the permuted data sets were always longer than the 30 steps associated with the original data. Likewise, permutation of the molecular data yields optimal trees that are much longer than the optimal tree for the original data, whose length was 790 steps (Figure 9.5). This finding corroborates the assertion that these data contain some information about the phylogenetic history of these organisms. In practice, it is extremely unusual to find data that completely lack hierarchical structure. Nonetheless, it is wise to be open to the possibility of a lack of treelike ancestry and, more generally, to understand how descent from common ancestry shapes the pattern of trait variation among living taxa.

MEASURES OF CLADE SUPPORT

Even if a data set contains phylogenetic signal, that does not mean that it provides equally strong support for all clades on the estimated trees. Ideally, we need some way to annotate the clades to indicate their relative strength of support from the data.

A simple measure of support for a clade is to look at the difference between the score of the optimal tree and the optimal tree that lacks that clade. In the case of parsimony, this difference is called the ***decay index***. It is also sometimes

called Bremer support in reference to Kåre Bremer, who promoted its use (Bremer 1994). In the case of likelihood, the statistic of choice is the likelihood ratio, that is, the difference in the log-likelihood scores of the optimal trees with and without this clade. For example, suppose that the optimal trees have length 292 and a log-likelihood of −998.7 and that all those trees have clade X. If the optimal tree lacking clade X has a length of 294 and a likelihood score of −1001.0, then the decay index of clade X would be 2 and the likelihood ratio would be 2.3. If clade Y has a decay index of +10 and a likelihood ratio of 12.2 for these same data, we could say that these data more strongly support clade Y than they support clade X. This is because the best tree without clade Y is much less consistent with the data than is the best tree without clade X.

The decay index and likelihood ratio can be readily understood with reference to tree space (Chapter 7). Recall that each region of tree space corresponds to a tree with a definite topology. The tree at a given point either has or does not have clade X. With reference to clade X, we can imagine tree space being divided into two regions—one composed of trees with clade X and one composed of trees that lack this clade. As shown in Figure 9.6, the decay index reports the difference between the highest peak in each of these two areas. Similarly, the likelihood ratio reports the difference in their log-likelihood scores, which equals the ratio of their raw likelihoods.

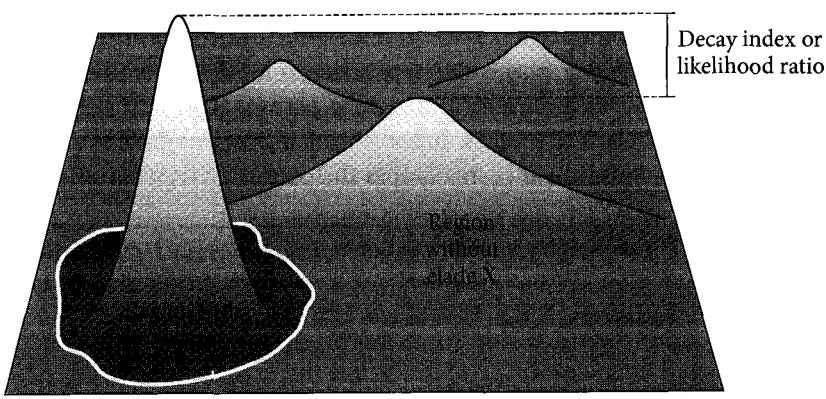


FIGURE 9.6 A tree space depiction of the decay index or likelihood ratio. The globally optimal tree lies in the region of tree space that includes clade X. The optimal tree lacking clade X (the highest peak outside the region of tree space that includes clade X) is identified. The difference in the tree length between that tree and the globally optimal tree is the decay index of clade X. The ratio of likelihoods provides an alternative measure of how strongly the data support clade X.

Decay indices and likelihood ratios are helpful measures of clade support. However, their magnitude will be affected by the number of characters and character states. As a result, further analyses are needed before we can conclude that a decay index or likelihood ratio is so great that we should reject the possibility that a tree lacking that clade is true.

The approaches that are used most commonly to evaluate clade support are called ***topology tests***. These tests compare the fit of two trees, differing in one or more clades, to the same data set. As an example, the parsimony-based ***Templeton test*** (developed by Templeton, 1983) assesses whether the number of characters that are shorter by a certain amount on each of the competing trees deviates from what we would expect if the characters had been equally likely to support each tree. If, using the Templeton test or another topology test, we conclude that the data are explained *significantly* better by the optimal trees that have clade X than by the optimal trees that lack clade X, then we would have more reason to believe that clade X is true. More information about topology tests is provided in the recommended Further Reading.

NONPARAMETRIC BOOTSTRAP ANALYSIS

The ***nonparametric bootstrap***, or just ***bootstrap***, has been the most widely used method for evaluating the strength of clade support since its use in phylogenetics was proposed by Felsenstein (1985a). Indeed, as mentioned in Chapter 3, a majority of phylogenetic tree diagrams in scientific publications have bootstrap scores attached to their branches.

Bootstrapping attempts to assess our chances of recovering a particular clade again if we were able to sample a new set of characters. The idea is that the data we have in hand are probably somewhat like the underlying universe of possible data sets. Therefore, we can simulate other possible data sets by drawing randomly from the data that we have already collected. Because this seems to get information from nothing, it has been equated with lifting yourself off the ground by your bootstraps; hence the procedure's name.

To illustrate the concept of a bootstrap, imagine the following (rather contrived) situation. We know that a con artist has been messing with a pair of decks by moving high-value cards (counting aces as low) from one "depleted" deck to another "enriched" deck, and moving low-value cards in the other direction. We do not know how many cards were moved or their values, but we know that the average value of a card drawn from the "enriched" deck

(counting jacks as 11, queens as 12, and kings as 13) is above 7, while that from the depleted deck is less than 7.

Imagine that we have managed to steal 10 cards from one of the decks and we want to assess whether the cards came from the depleted or enriched deck. If the average value of these 10 cards is greater than 7, we might suspect that they came from enriched deck. If their average value is less than 7, we might, instead, suspect that we had drawn them from the depleted deck. The problem is that even if the cards came from the depleted deck, we might have unluckily drawn high cards, giving us the impression that the deck was enriched (and vice versa). A nonparametric bootstrap provides a way to assess this possibility without having to steal another 10 cards.

Before explaining the bootstrap method, let us clarify how this scenario relates to the problem of phylogenetic inference. There is some true tree, which either has or lacks a certain clade, X. If clade X is true, we expect to see an excess of characters that are consistent with clade X. This is analogous to having an excess of high-value cards in the enriched deck. Conversely, if clade X is false, we expect a deficit of characters that support clade X, analogous to the depleted deck. Drawing ten cards from the deck and asking whether the deck was enriched or depleted is analogous to collecting data from a set of taxa and asking whether clade X is true or false. As with the cards, we could draw a data set that supports clade X even though clade X is false. The nonparametric bootstrap provides a way to assess how likely it is that our conclusions are sound.

In the card case, a nonparametric bootstrap would proceed as follows. We randomly draw one of the 10 cards and note its value. We replace that card in the pile of 10, shuffle, and draw a second card. After noting its value, we repeat the exercise, and keep doing the same until we have noted the value of 10 cards. The 10 values we have drawn randomly, with replacement, from the original 10 cards is a bootstrap sample. It is also called a ***pseudoreplicate*** because it resembles our original data set of 10 cards, yet is not really a replicate in that we did not collect any new data. We would note the average value of the 10 cards in the bootstrap sample and then create many more bootstrap samples (typically at least 100), noting the average card value of each pseudoreplicate. If the great majority, for example, 95%, of the bootstrap samples are composed of cards with an average value over 7, then we would have good reason to conclude that the cards did indeed come from the enriched deck. If, however, many of our pseudoreplicates have average card values that are less than 7, we should conclude that such a sample of 10 cards is insufficient to be confident that the deck was enriched.

You may be wondering how the number of bootstrap samples with an average value >7 relates to the probability that the true deck has an average card value >7 (i.e., that it is enriched). It turns out that, if the set of 10 cards were drawn independently and randomly from the deck, then, even if their mean value deviated from the true mean for the deck, there is a high probability that the variance in card value would be similar to the variance in that deck. By sampling with replacement from our hand, we are using the variance in our sample to provide an estimate of how confident we should be in conclusions drawn from the original set of 10 cards.

The application of nonparametric bootstrapping is similar. Imagine that we have collected data, which yielded a single optimal tree (by whatever method) that has clade X, and we want to evaluate how confident we should be in clade X. We cannot collect additional data, so instead we generate a large number of bootstrap data sets. Each of these pseudoreplicates includes the same number of characters (i.e., columns) as the original data, and is assembled by sampling with replacement from the original data. Table 9.2 shows a small data set for eight taxa (A–H) and one bootstrap replicate. Note that some characters from the original data (e.g., character number 6) are sampled more than once, whereas others (e.g., character number 4) are not sampled at all.

Because the bootstrap data set contains a different sampling of characters, it could yield a different optimal tree than the original data. We can take the bootstrap data set and analyze it using the same method as we used for the original data (e.g., maximum parsimony, neighbor-joining, maximum likelihood). This is analogous to determining the average card value for a bootstrap hand. For example, the single most parsimonious trees (with taxon A defined as the outgroup) from the original and bootstrapped data in Table 9.2 are shown in Figure 9.7. These two trees are different, but nonetheless they share three clades: (D,E,F,G,H), (F,G,H), and (G,H).

In a full bootstrap analysis, the proportion of the bootstrap data sets that yield a tree with a given clade is recorded. This is the ***bootstrap score*** or ***bootstrap percentage*** for that clade. For example, if 750 of 1000 bootstrap data sets yield an optimal tree with clade X, clade X has a bootstrap score of 75%. It is common to summarize these results by showing a ***bootstrap consensus tree*** (sometimes, loosely, called a ***bootstrap tree***), which is composed of those clades with the highest bootstrap scores. Alternatively, clades that appeared on the optimal tree can be annotated with their corresponding bootstrap scores. For the data in Table 9.2, the bootstrap consensus tree obtained using parsimony is shown in Figure 9.8. The numbers on an internal branch show the

TABLE 9.2 Generating a bootstrap replicate from a set of observed data

Original data set

	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
A	T	T	T	C	C	T	T	T	C	A	G	G	T	A	T	T	A	T	G	A	G	A	T	A	C	G	T	A	C	T	G	A	A	A	A	A	G	T	C	C
B	T	T	T	C	C	T	T	T	T	A	G	G	T	T	T	G	A	T	G	A	G	A	T	A	C	A	T	T	A	C	G	A	A	A	G	A	G	T	C	A
C	T	T	T	G	C	T	T	C	T	C	G	G	T	A	C	T	A	C	A	A	T	A	T	A	T	A	T	A	C	C	A	G	A	A	A	A	G	T	C	A
D	T	T	T	G	C	T	T	C	C	G	A	C	T	A	C	A	A	A	G	G	C	A	T	A	C	G	T	A	G	C	T	G	A	A	A	A	G	G	C	G
E	C	T	T	G	C	C	T	A	C	T	G	T	T	G	C	A	A	T	A	A	T	A	T	A	C	G	A	A	G	C	T	A	A	A	A	A	G	T	C	G
F	T	T	C	G	T	C	C	C	C	G	G	C	T	A	C	A	A	T	G	G	T	A	T	A	T	G	T	A	C	T	C	G	A	A	A	A	G	A	T	G
G	G	T	T	G	T	T	T	C	C	G	G	C	T	A	C	A	G	T	G	A	T	A	T	A	C	G	T	A	C	C	C	G	A	G	A	A	C	T	T	G
H	T	T	T	A	T	T	T	C	C	G	G	C	T	A	C	A	G	T	G	A	T	A	T	A	C	G	T	G	C	C	C	G	A	G	A	A	G	T	T	G

Bootstrap data set

	02	39	35	22	36	31	40	05	16	23	15	35	35	40	03	06	24	33	06	07	14	20	35	01	36	09	13	22	11	25	26	33	03	09	16	20	08	18	17	32
A	T	C	A	A	A	G	C	C	T	T	T	A	A	C	T	T	A	A	T	T	A	A	A	T	A	C	T	A	G	C	G	A	T	C	T	A	T	T	A	A
B	T	C	G	A	A	G	A	C	G	T	T	G	G	A	T	T	A	A	T	T	T	A	G	T	A	T	T	A	G	C	A	A	T	T	G	A	T	T	A	A
C	T	C	A	A	A	A	A	C	T	T	C	A	A	A	T	T	A	A	T	T	A	A	A	T	A	T	T	A	G	T	A	A	T	T	T	A	C	C	A	G
D	T	C	A	A	A	T	G	C	A	T	C	A	A	G	T	T	A	A	T	T	A	G	A	T	A	C	T	A	A	C	G	A	T	C	A	G	C	A	A	G
E	T	C	A	A	A	T	G	C	A	T	C	A	A	G	T	C	A	A	C	T	G	A	A	C	A	C	T	A	G	C	G	A	T	C	A	A	A	T	A	A
F	T	T	A	A	A	C	G	T	A	T	C	A	A	G	C	C	A	A	C	C	A	G	A	T	A	C	T	A	G	T	G	A	C	C	A	G	C	T	A	G
G	T	T	A	A	A	C	G	T	A	T	C	A	A	G	T	T	A	A	T	T	A	A	A	G	A	C	T	A	G	C	G	A	T	C	A	A	C	T	G	G
H	T	T	A	A	A	C	G	T	A	T	C	A	A	G	T	T	A	A	T	T	A	A	A	T	A	C	T	A	G	C	G	A	T	C	A	A	C	T	G	G

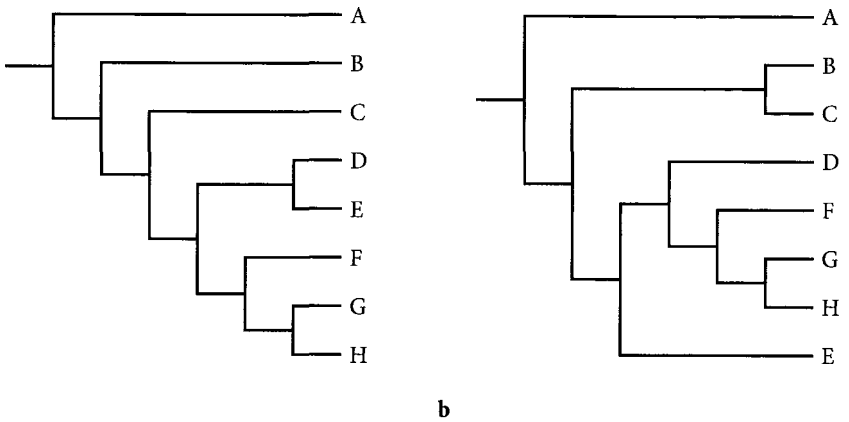


FIGURE 9.7 Trees built from original data set (a) and bootstrap data set (b) in Table 9.2. Taxon A is treated as an outgroup for the purposes of rooting.

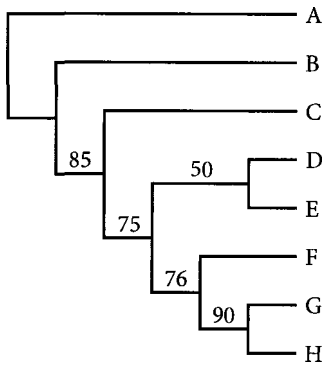
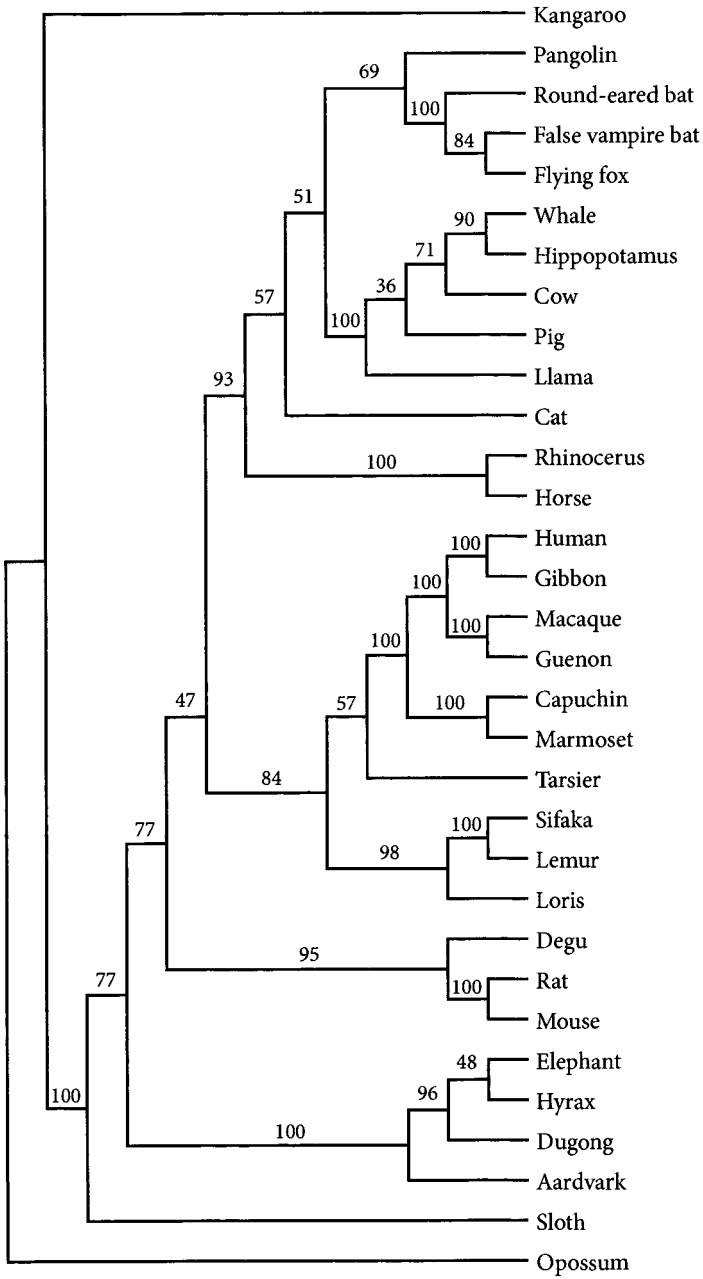


FIGURE 9.8 Parsimony bootstrap consensus tree for the data in Table 9.2. The bootstrap scores associated with each internal branch indicate the average percentage of most parsimonious trees that have the clade among 100 bootstrap pseudoreplicates.

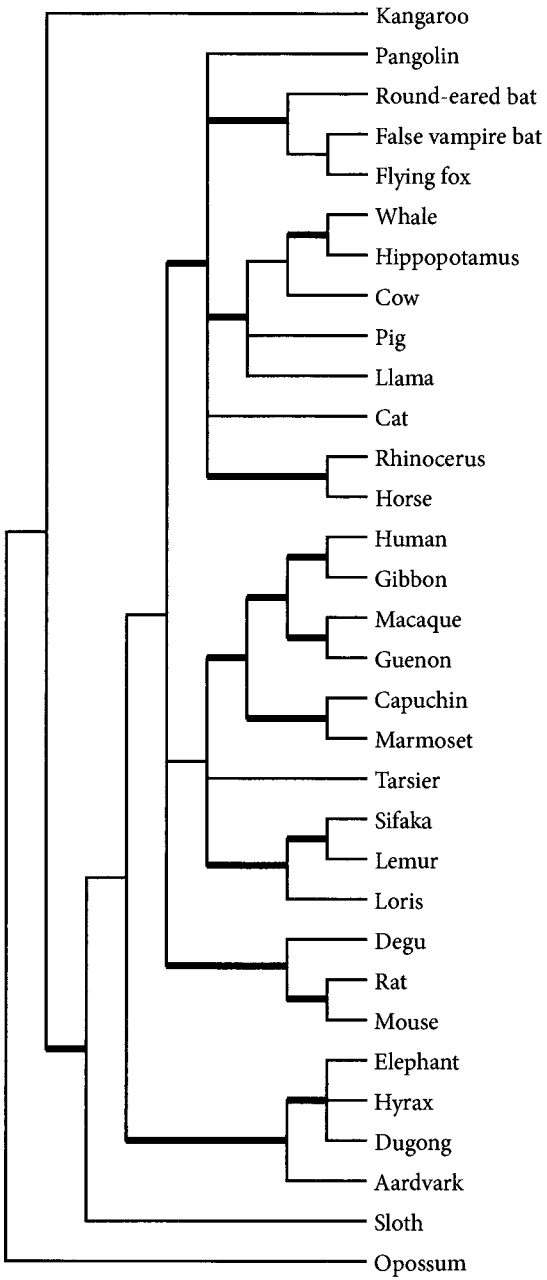
proportion of bootstrap samples that have the clade. On the basis of this result, we could conclude that the (G, H) clade, with a bootstrap score of 90%, is supported more strongly than is the (D, E) clade, with a bootstrap score of 50%.

As an example from the literature, Figure 9.9 shows a parsimony bootstrap consensus tree from an analysis of three genes sequenced for several placental



a

FIGURE 9.9 Collapsing branches that are only weakly supported as applied to a phylogeny of placental mammals. In (a) bootstrap scores are indicated as numbers on branches. In (b) branches with less than 70% bootstrap are collapsed and branches having scores over 90% are drawn extra wide. The original data (from which a subsample of taxa was analyzed here) comes from Bininda-Emonds et al. (2006). (Figure continues on next page.)


Kangaroo
Pangolin
Round-eared bat
False vampire bat
Flying fox
Whale
Hippopotamus
Cow
Pig
Llama
Cat
Rhinocerus
Horse
Human
Gibbon
Macaque
Guenon
Capuchin
Marmoset
Tarsier
Sifaka
Lemur
Loris
Degu
Rat
Mouse
Elephant
Hyrax
Dugong
Aardvark
Sloth
Opossum

b

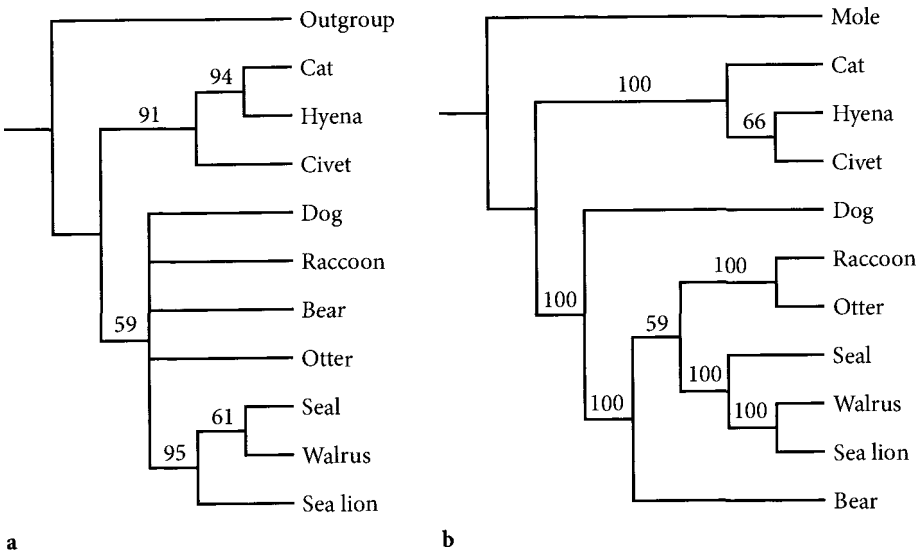


FIGURE 9.10 Bootstrap consensus trees based on the carnivoran morphological (a) and molecular (b) data. Clades with bootstrap scores less than 50% have been collapsed.

mammals (rooted with a marsupial). You will see a range of bootstrap scores. To make it easier to see the well-supported conclusions, the second tree reports the same conclusions, but with all branches having bootstrap scores less than 70% collapsed and branches having scores over 90% drawn extra wide.

Figure 9.10 shows the parsimony bootstrap consensus trees for the carnivoran morphological and molecular data. You can see that, even from these rather small data sets, several clades appear to gain strong statistical support. The morphological data yield fewer clades with high bootstrap scores than do the molecular data, which probably reflects the lower number of characters in the morphological data matrix. While the molecular data indicate strong support for most relationships, two clades remain uncertain. It is noteworthy that there are no cases where a clade that is well supported by one of the two data sets contradicts a well-supported clade from the other data set. This is consistent with the hypothesis that the two data sets arose from the same underlying phylogeny.

PARAMETRIC BOOTSTRAPPING

Nonparametric bootstrapping generates new data sets by sampling with replacement from the original data. Parametric bootstrapping, in contrast, generates new data sets by simulating them using an evolutionary model. Whereas nonparametric bootstrapping is applicable to many different methods of phylogenetic analysis, parametric bootstrapping is primarily used in conjunction with maximum likelihood analysis.

Imagine that we have found the maximum likelihood estimate of the tree topology, branch lengths, and model parameters. Given this set of parameters we can ask, What data set might plausibly have evolved instead of the one we actually obtained? This is usually done using computer simulation, for example, using the program Seq-Gen (Rambaut and Grassly 1997). A random sequence conforming to the model's assumptions (e.g., equality of base composition) is generated at the base of the tree and is then allowed to evolve along the branches of the tree. For each branch we can use the substitution probability matrix and the estimated model parameters (e.g., the transition:transversion bias) to determine the probability that a base starting in one state at the ancestral node will be found in each of the four possible states at the descendant node. Using a random-number generator, we can let chance decide which of the outcomes actually happens in our simulation. By repeating the process for all positions in the sequence and for all branches of the tree, we can generate a new data matrix. As with nonparametric bootstrapping, we can repeat this process hundreds or thousands of times and examine how conclusions vary across these pseudoreplicates.

You may be wondering why one would use parametric bootstrapping, since it does something very similar to nonparametric bootstrapping but is much more laborious. For evaluating clade support, there is no great advantage to parametric over nonparametric bootstrapping. Both methods provide a sense of the degree to which clade support is robust to sampling error. Both do so by assuming that the actual data set is representative of the universe of data sets that might have evolved. Thus, the less intensive approach of nonparametric bootstrapping is typically used for assessing clade support.

In contrast, parametric bootstrapping is often used to test specific phylogenetic hypotheses. As an example, we will consider the application of parametric bootstrapping to address the controversial phylogenetic placement of

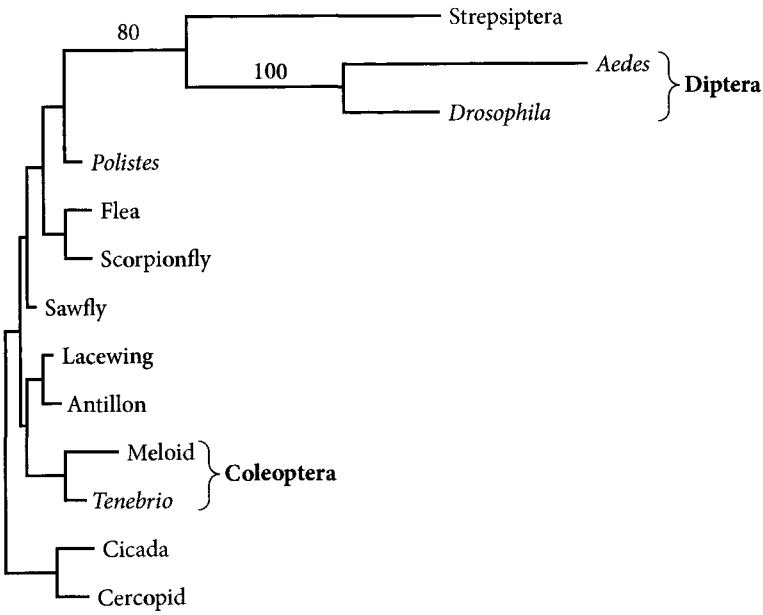


FIGURE 9. 11 Inferred phylogeny of insects with Strepsiptera sister to Diptera. The bootstrap support for two clades is shown. Adapted from Huelsenbeck (1997).

an enigmatic group of parasitic insects, the Strepsiptera. A parsimony analysis by Whiting et al. (1997) suggested that Strepsiptera are the sister group to the true flies (Diptera, including *Aedes* and *Drosophila*). This result conflicts with the traditional view that Strepsiptera are sister to the beetles (Coleoptera, including Meloid and *Tenebrio* beetles). Huelsenbeck (1997) reviewed the result and noticed that Diptera and Strepsiptera shared long branches (Figure 9.11), and wondered if long-branch attraction (Chapter 7) rather than actual evolutionary kinship could be responsible for the topology obtained in Whiting et al. (1997).

To explore this question further, Huelsenbeck created a ***constraint tree***, a tree that is forced to have a particular clade or topology. In this case, the constraint placed Strepsiptera sister to Coleoptera, following the traditional view. He then found the optimal tree under this constraint (Figure 9.12) and simulated data up this tree numerous times. Finally, he took the resulting data sets and determined the optimal parsimony tree for each. Interestingly, for 92% of

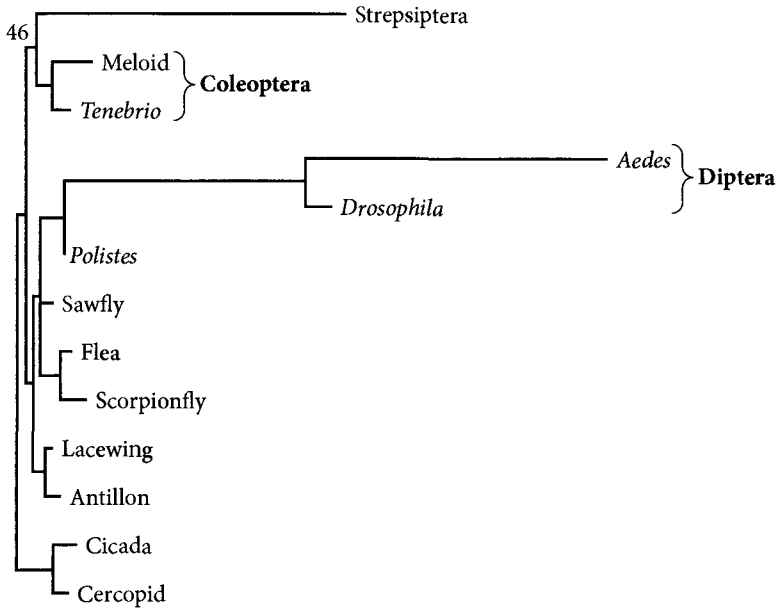


FIGURE 9.12 Constraint tree with Strepsiptera sister to beetles. This is the optimal tree found that includes a Coleoptera + Strepsiptera clade. Adapted from Huelsenbeck (1997).

the data sets, Diptera and Strepsiptera appeared as sister groups, despite the fact that the data had been simulated along a tree that had Strepsiptera sister to Coleoptera. This result shows that branch-to-branch rate heterogeneity could easily cause data to evolve that supported the incorrect conclusion that Strepsiptera and Diptera are sister groups even when the true evolutionary history included a Strepsiptera + Coleoptera clade.

This finding did not prove that Strepsiptera *are* the sister group to the beetles. It just shows that the data analyzed by Whiting et al. (1991) were not definitive on this point. The parametric bootstrap analysis, thus, helped to avoid erroneously rejecting the Strepsiptera + Coleoptera hypothesis based on this single set of data. Subsequent research (e.g., Longhorn et al. 2010) employing more molecular data (over 10,000 bases) and greater taxon sampling has now supported the traditional view of a sister relationship of Strepsiptera and Coleoptera.

BAYESIAN STATISTICAL ANALYSIS

All of the statistical analyses described so far in this chapter are applicable to methods that yield a single or small set of optimal trees, for example, parsimony or maximum likelihood. These methods yield a ***point estimate*** of that which we wish to estimate, the phylogeny. As a result, we need to conduct additional analyses to obtain a confidence interval around this tree estimate.

Bayesian analysis is fundamentally different because what we obtain is not a point estimate but something that already contains information on statistical uncertainty: a posterior distribution. This distribution contains a sample of trees that could plausibly have given rise to the data, ranked by the probability that each is the true tree (given the data, model, and prior distribution; see Chapter 8). By exploring this distribution, we can address the same questions we aimed to answer with measures like decay indices and bootstrap scores.

We can examine the output of a Bayesian Markov chain Monte Carlo (MCMC) analysis to determine the posterior probability that a particular phylogenetic hypothesis is true. For example, having done a Bayesian analysis of the data underlying Figure 9.11, we could search through the trees sampled from the posterior distribution and determine the proportion of trees that are compatible with the hypothesis that Strepsiptera are sister to Coleoptera. If only 1% of trees were compatible with this prior hypothesis, then we ought to be able to state that there is a 99% probability that the hypothesis is false. Because Bayesian inference yields a full statistical distribution, we can "test" hypotheses by simply querying the Bayesian posterior distribution to assess how much of it is or is not compatible with different hypotheses. However, a few caveats should be noted.

At the end of a Bayesian MCMC analysis, one has a computer file that lists a few thousand trees that are sampled from the posterior distribution. Assuming that the MCMC analysis mixed properly, the frequency of trees in this file is an estimate of their posterior probability. But some error is expected. For example, many trees are not represented at all in the posterior distribution, even though all trees have a nonzero posterior probability. Thus, the frequency of a tree topology or a clade is expected to deviate somewhat from its true posterior probability. On the other hand, the magnitude of these errors will usually be small, especially when the MCMC sample contains many trees.

A much bigger problem is that the posterior distribution is valid given a specific prior distribution and model of molecular evolution. Thus, the accuracy of the posterior distribution depends on how well the model describes the process that generated the data and how well the prior distribution captures our prior knowledge. If we are testing a prior hypothesis about phylogenetic relationships, then we likely have evidence that trees compatible with this hypothesis have higher prior probability than trees that are incompatible with the hypothesis. For example, in the Strepsiptera case, there was prior evidence that Strepsiptera and Coleoptera are sister taxa because of shared larval traits. However, standard phylogenetic MCMC methods are not able to incorporate different prior probabilities for different topologies. Instead, Bayesian inference almost always uses "flat" or "uninformative" prior probabilities on different topologies, implying that all trees (or all trees within a subset of tree space being considered) are equally likely, *a priori*. If an MCMC analysis does not incorporate the true prior knowledge on tree topology, we should think twice about rejecting a phylogenetic hypothesis even if its posterior probability is extremely low.

Several methods, such as Bayes' factor analysis, are available to help one decide on the relative posterior probabilities of two competing phylogenetic hypotheses even in the case where a Bayesian MCMC analysis was conducted using uninformative prior probabilities. The recommended Further Reading provides additional information on these statistical approaches.

COMPARING DATA SETS

Beyond testing whether a single data set supports or refutes a particular phylogenetic hypothesis, we often need statistical methods to compare multiple data sets and determine if they share the same underlying phylogenetic history. As we saw with the carnivoran data set, we can draw on different sources of data to estimate phylogenies for the same set of taxa, and these data can give different trees. These different sources of data are called ***partitions*** because they could potentially be combined into a single data set. The carnivoran data had two partitions: DNA sequences and morphology.

Even when only one kind of data has been collected, for example, DNA sequences, the data set could be partitioned. Different genes could be treated as different partitions and partitions could be defined within a single gene based

on coding status (e.g., introns vs. exons; first- or second- vs. third codon positions) or functional constraints (e.g., extracellular vs. transmembrane protein domains). Generally, a partition is a subset of data that we believe might have a different history or a different model of evolution from other partitions.

In most cases, we expect different partitions sampled from the same taxa to yield identical or very similar trees. This fact has been highlighted as providing evidence for evolution (Chapter 2). However, while agreement among partitions is usually only explicable under an assumption of evolution from common ancestry, a lack of agreement among partitions does not refute common ancestry. There are a number of well-understood phenomena that cause different data sets from the same organisms to track different phylogenetic histories, including introgression and incomplete lineage sorting (Chapter 6). Determining whether two (or more) partitions derive from distinct evolutionary histories is a first step in exploring these interesting biological phenomena.

In addition to the biological reasons for assessing conflict between partitions, there are methodological considerations. If two partitions share the same history, we can assemble them into a single ***combined*** (or ***concatenated***) data matrix. Because this matrix will have more characters, it is expected to have more phylogenetic signal and to provide a better phylogenetic estimate of the true tree. If, however, the two partitions have different histories, combining them into a single matrix would not be expected to provide clear evidence about either partition.

Figure 9.13 shows a hypothetical data matrix with two partitions of 40 DNA sequence characters. As shown in Figure 9.14, there is just one most parsimonious tree for partition 1 and three equally parsimonious trees for partition 2, whose consensus is shown. The optimal trees for the two partitions differ, for example, in the presence of a (D, E, F) clade. This does not mean that the

	Partition 1	Partition 2
A	TTTAGATCTCACAATTTCGTGGGCAACATCACTTGCCAGA	GACGTAATCACCCAAGCCCGTTGCCTCCGGAAACGACGTG
B	TTTAGATTTCACAAGTTCGTGGGCAACGTCACTTACCAGA	GACGTAGGTACCCAAGCCCCTTGCCTTGGTAAACGACGTG
C	TTTAGATTTCACAAGTTCGAGGGGAGCACGACTTGTCAAA	GACGTAATCACCTAAGTCCCTTGCGTCCGTAAACGACGCG
D	TTCATACCCTACGAGGTCATGGGCATCACGACTTATCAGA	GACGCAACAATTAAAGCCCTTTGTACGCGGAAACCGCGTG
E	TTCAAATTTGACGACTTCGTGGGCATGACGACTTATCAGT	GACGCAACAACTAAACCCCTTTGTACGCGGAAACCACGCA
F	TTCGGGCTTTAGTAGTCCCTGGGCAGCACAATTAGTCGTA	AACGTAACAACTGAACCCCCTTGTTCGCGTAGACTGCATG
G	TTTAAGTCTCAGGAATCGCTAGGCAGCACAATTTGTCTTA	GACGTAACCACTAAACCCCCCGGTCTCCGCTGATGACGTG
H	TTCAGGTTTCAAGAATCGTTGGGCAGCACAATTTGTCCTA	GACCTAACCACTAAGCCCCCCGGTCTGCGCGGATGACGTG

FIGURE 9.13 Two partitions of a hypothetical molecular data set.

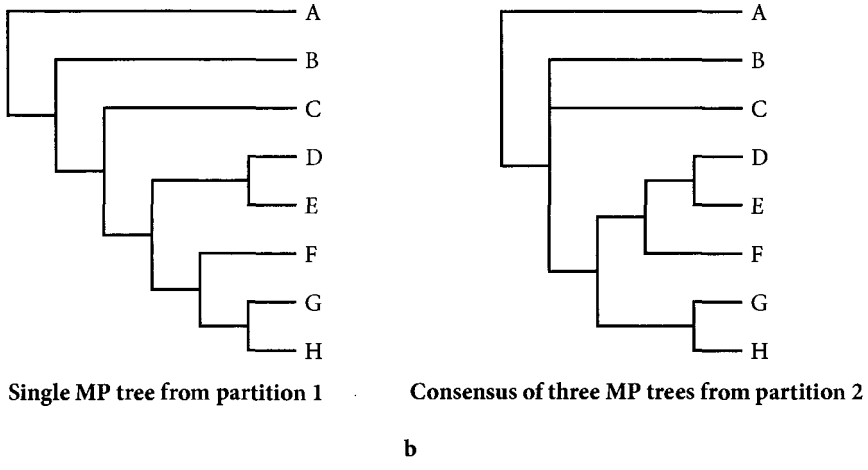


FIGURE 9.14 **Most parsimonious trees from each partition of Figure. 9.13.** (a) Single most parsimonious tree from partition 1. (b) Consensus tree of the three equally parsimonious trees from partition 2. While the two trees are similar, they differ in the placement of taxon F.

partitions have necessarily tracked different histories. It could be that they have tracked the same history, but chance has resulted in a misleading tree from one or both data sets. How can we assess whether partitions derive from the same tree?

One strategy for assessing data set conflict is to consider if the optimal tree from one partition (or from the combined data set) is consistent with the other partitions. This can be assessed with a topology test. To remind you, a topology test asks whether the optimal tree for a data set is so much better than an alternative tree that it is improbable that the data derive from that alternative tree. A Templeton test, for example, could be used to determine if a partition is significantly more compatible with its own optimal tree than with another partition's optimal tree. If the answer is negative (the alternative tree is no worse than the optimal tree), then we would have reason to conclude that the data sets all arose from the same tree.

In the case of the data in Figure 9.13, a Templeton test shows a lack of conflict. Although partition 1 tends to reject the trees from partition 2 (P-values of 0.01–0.03, depending on which of the three equally parsimonious trees are considered), the data from partition 2 cannot reject the optimal tree from partition

1 ($P = 0.56$). This result is gratifying because, in fact, both partitions were simulated up a single tree (similar to the tree for partition 1).

In the case of the carnivoran data, a Templeton test shows that the trees for the morphology data are rejected by the molecular data ($P < 0.0001$). However, the most parsimonious trees for the molecular data (one of which also happens to be the optimal tree for the combined data) are *not* significantly rejected by the morphological data ($P = 0.08 - 0.18$). This means that both the molecular and morphological data sets could plausibly have evolved along the tree topology supported by the molecular data. This indicates that, despite the differences between their optimal trees, the two data sets are not in conflict.

As a note of caution, if you fail to find a tree that is compatible with all partitions, you should not conclude that partitions conflict. There could be another tree that you have not considered that could have generated all the data. The presence of conflict can only be confirmed by topology tests if you find a specific clade that is significantly supported by one partition (i.e., the optimal trees lacking the clade are rejected) and significantly rejected by the other partition (i.e., the optimal trees having the clade are rejected).

PARTITION HOMOGENEITY

An alternative approach to evaluating data set conflict is to ask whether partitions appear to be drawn randomly from the same underlying population of characters. This may be studied with the ***partition homogeneity test***, also called the ***incongruence length difference test*** (Farris et al. 1994). This method randomly assigns characters to partitions many times and then conducts a phylogenetic analysis of the random partitions. This analysis can use parsimony, minimum evolution, or maximum likelihood, but we will describe the approach using parsimony.

As discussed in relation to the PTP test, as we decrease the phylogenetic signal within a data set, we tend to increase the length of the optimal tree because of conflict among characters. The same will happen if we combine data partitions that have conflicting signals. If two partitions have tracked different trees, we expect the sum of the lengths of the most parsimonious trees for each partition to be less than the sum of the lengths of the trees obtained from two random partitions that each contain a mixture of characters from the original partitions. This will be easier to understand by working through the actual procedure.

Partition 1

```
A TTCGAGAACACGGCCCTTTGCGACCCATGTTGTTA
B TCAGAGAACACGACACTTTGCGACCCATGTTGTTA
C TCCGAGAGCACGGACCTTCGCGACCTATGTTATTG
D TCCGGAAGTAAACACCCTTGTAACCCCAGCTATCG
E TCCGGAAGTAGACGCCCTTGTAACCCCAGCTATCG
F TCCGGGAGTAAATGCCCTTGGGACCCCTGCTATTG
G TCTGGGAGCACAAGTCCTCACGACCCCTGCTATTG
H TCTGGGAGCACAAATCCTCACGACCCCTGCTATTG
```

Partition 2

```
A TTCGAGAACACGGCCCTTTGCGACCCATGTTGTTA
B TCCGGAAGTAAACACCCTTGTAACCCCAGCTATCG
C TCCGGAAGTAGACGCCCTTGTAACCCCAGCTATCG
D TCCGAGAGCACGGACCTTCGCGACCTATGTTATTG
E TCTGGGAGCACAAGTCCTCACGACCCCTGCTATTG
F TCTGGGAGCACAAATCCTCACGACCCCTGCTATTG
G TCAGAGAACACGACACTTTGCGACCCATGTTGTTA
H TCCGGGAGTAAATGCCCTTGGGACCCCTGCTATTG
```

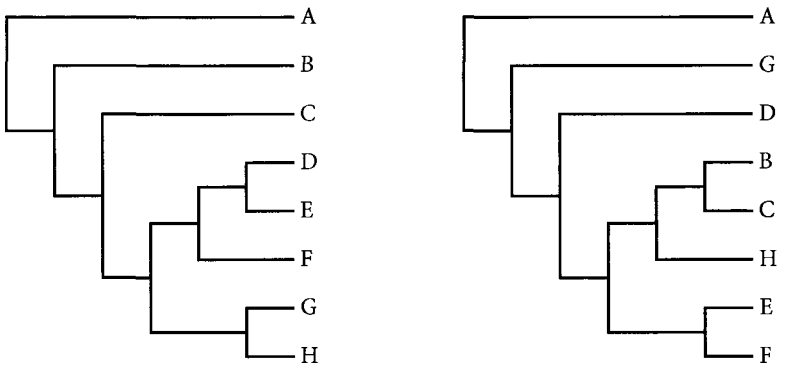


FIGURE 9.15 An example of the partition homogeneity test. The two partitions were artificially generated in order to be in conflict. The sets are identical except that the sequences have been assigned to different taxa. Their respective optimal trees are shown below.

We have generated two data sets with conflicting phylogenetic signals (Figure 9.15), allowing us to see if the partition homogeneity test successfully detects this conflict. We created this artificial case by taking the first data set (partition 1) and swapping the sequences among taxa to generate a second data set (partition 2). While the two partitions support a different set of relationships, the lengths of their respective optimal trees are identical (34 steps), as would be expected given that they contain the same sequences.

If we generate a composite data set with half of partition 1 and half of partition 2, the shortest tree is either 41 or 44 steps (depending upon which half of each data set is included). The reason that these trees are longer than the original data sets is that they combine data with conflicting phylogenetic signals. If two data sets have a different signal, then generating composite data sets will tend to increase the length of the trees (and decrease the likelihood score).

```
                                                                                  Partition
                                                                                  assignment
  12212122111212212112212121121221212211121212212122121112121121221122 1
A TTCGAGAACACGGCCCTTTGCGACCCATGTTGTTATTCGAGAACACGGCCCTTTGCGACCCATGTTGTTA
B TCAGAGAACACGACACTTTGCGACCCATGTTGTTATCCGGAAGTAAACACCCTTGTAACCCCAGCTATCG
C TCCGAGAGCACGGACCTTCGCGACCTATGTTATTGTCCGGAAGTAGACGCCCTTGTAACCCCAGCTATCG
D TCCGGAAGTAAACACCCTTGTAACCCCAGCTATCGTCCGAGAGCACGGACCTTCGCGACCTATGTTATTG
E TCCGGAAGTAGACGCCCTTGTAACCCCAGCTATCGTCTGGGAGCACAAGTCCTCACGACCCCTGCTATTG
F TCCGGGAGTAAATGCCCTTGGGACCCCTGCTATTGTCTGGGAGCACAAATCCTCACGACCCCTGCTATTG
G TCTGGGAGCACAAGTCCTCACGACCCCTGCTATTGTCAGAGAACACGACACTTTGCGACCCATGTTGTTA
H TCTGGGAGCACAAATCCTCACGACCCCTGCTATTGTCCGGGAGTAAATGCCCTTGGGACCCCTGCTATTG
```

FIGURE 9.16 Random assignment of positions to partitions in the partition homogeneity test. Data from the upper half of Figure 9.15.

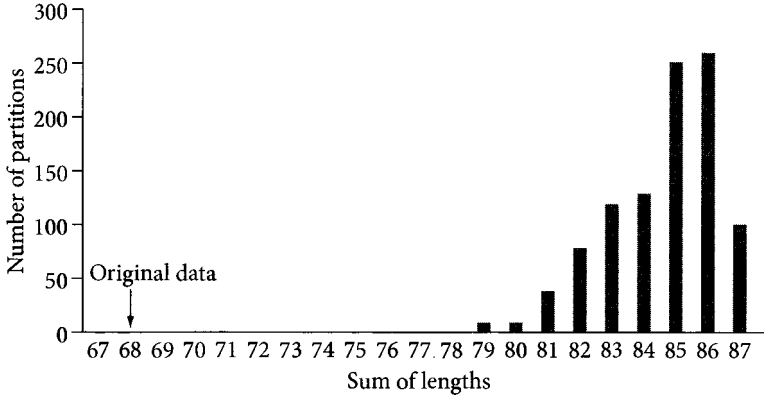


FIGURE 9.17 Sum of tree lengths based on 100 random partitions of the data shown in Figure 9.15. The summed tree lengths of the original data sets is 68. Because the random partitions yield trees that are all much longer than the original data, we can conclude that there is significant conflict between the two partitions.

To do a partition homogeneity test, we first determine the length of the most parsimonious trees for each partition and sum up the length across partitions. In this example, the two optimal trees sum to 68 steps. Next, we randomly reassign characters to the two partitions. For example, Figure 9.16 shows a random assignment of characters (columns) to two partitions, with 35 characters in each (as in the original data).

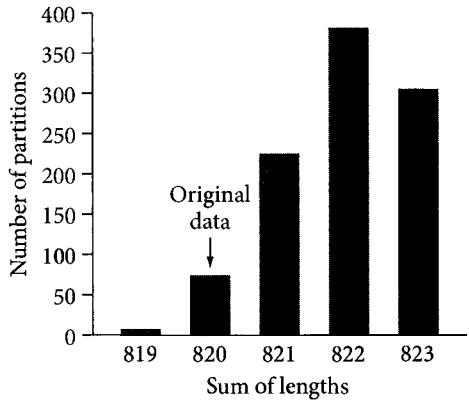


FIGURE 9.18 Sum of tree lengths for 1000 random partitions of the carnivoran molecular and morphological data. The summed lengths of the original data sets is 820. Because many of the random partitions have an equal or shorter sum of lengths, we can conclude that there is not significant conflict between the molecular and morphological data.

The shortest tree for random partition 1 has length 39, whereas the shortest for partition 2 has length 46, giving a sum of 85 steps. The partition homogeneity test repeats this procedure many times and sees how often the sum of the two partitions is the same or shorter than the original data. Figure 9.17 shows the distribution of lengths for 100 random partitions. Because none of the random partitions yields pairs of trees that are as short as 68 steps, we have good reason to conclude that the original partitions are nonrandom, implying that the partitions have distinct phylogenetic signals. This would argue that the two partitions do not share the same phylogenetic history. Given the obvious conflict between these two data partitions, they should not be combined into a single data matrix.

Figure 9.18 shows the result of a partition homogeneity test applied to the carnivoran morphological and molecular data sets. The sum of the shortest tree for these two data sets is 790 + 30 = 820. Out of 1000 randomized partitions, 85 have a sum of lengths that is equal (820) or lower (819) than the original data set. This corresponds to a P-value of 0.085. This can be interpreted loosely as meaning that, had these two data sets been randomly selected from the same universe of characters, there is an 8.5% chance that they would have this much (or more) conflict. Using a conventional cutoff for significance of

5%, this result agrees with the Templeton test in suggesting that the carnivorans' morphological and molecular data could have evolved along the same underlying phylogeny.

SPECIES TREE METHODS

The presence of different underlying histories for different data partitions is the expected result of a number of well-understood evolutionary processes (Chapter 6). For example, when one population lineage-splitting event quickly follows another, some genes may experience incomplete lineage sorting, causing those genes to have a true tree that is different from the population tree. Also, genes from the same population lineages can disagree if there has been gene flow via introgression.

In recent years, a new set of methods has been developed that allow for different (recombinational) genes to have potentially conflicting histories while estimating the treelike history of populations, commonly referred to as a ***species tree***. We will use this term despite the fact that it implies, incorrectly, that the tips must necessarily be ranked as species. Although the methods for ***species tree inference*** are quite sophisticated, we believe it is important to provide an introduction to them because of their growing importance in phylogenetic analysis.

Species tree methods have been developed within the parsimony, maximum likelihood, and Bayesian frameworks. Most assume that all gene-to-gene discordance is due to incomplete lineage sorting. If this is the case, then the true population history is treelike, which is to say lacking reticulation due to lateral gene transfer, introgression, or lineage fusion (Chapter 6). The challenge is to estimate the topology of this species tree given information in a set of gene trees.

The simplest approach is a parsimony-based method called ***minimizing deep coalescences***, or MDC (Maddison and Knowles 2006). As discussed in Chapter 6, incomplete lineage sorting requires that genes in two terminals coalesce at least one node deeper in the species tree than their last common ancestral population. Thus, incomplete lineage sorting needs to be invoked whenever we observe such deep coalescence. If we hypothesize a species tree, we can count up the minimum number of deep coalescence events needed to explain the topology of that gene tree. If we have many gene trees, we can

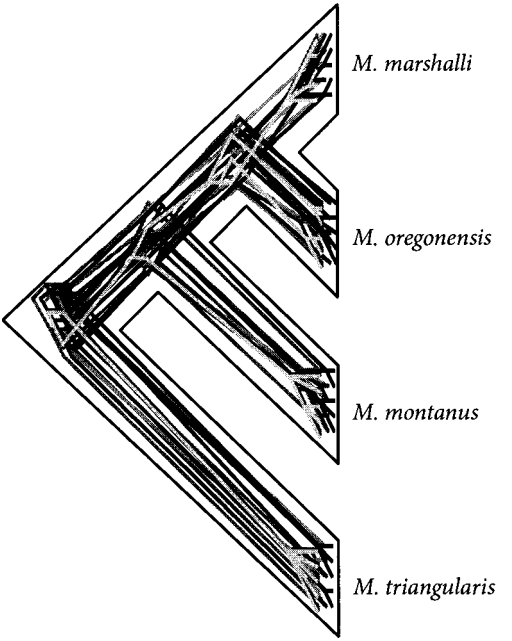


FIGURE 9.19 Species tree for four species of *Melanoplus* grasshoppers inferred using the minimizing deep coalescence approach. The different gene trees are "folded" into the population tree under the assumption that all gene-to-gene discordance is due to incomplete lineage sorting. Adapted from Carstens and Knowles (2007).

search among species trees to find the topology that minimizes the total number of deep coalescence events (summed over all gene trees). Based on coalescence theory, a branch of population genetics, this tree is a good estimate of the species tree.

Figure 9.19 provides an example of the application of the MDC method. Carstens and Knowles (2007) analyzed six genes from four closely related species of *Melanoplus* grasshoppers and concluded that the species tree shown invokes the lowest number of deep coalescence events and is the best estimate of the species tree.

Model-based approaches, whether implemented in a maximum likelihood or Bayesian framework, are natural extensions of the MDC approach. Despite being grounded in coalescence theory, MDC does not make good use

of coalescence models. For example, MDC only looks at tree topology, failing to use branch length estimates from the individual gene trees, even though this information can help us identify deep coalescence. Model-based methods search for the species tree topology and branch lengths that best explain the observed gene trees. A species tree's branches are measured in ***coalescence units***, the product of the population size and the number of generations, because this is what predicts the probability of incomplete lineage sorting. The more sophisticated, model-based species tree methods simultaneously take account of uncertainty in gene tree estimation and the chance aspects of incomplete lineage sorting. These are elegant methods, but often slow because of their computational complexity.

As an example of model-based species tree estimation, Figure 9.20 gives the results from analyses of North American bats in the genus *Myotis*, conducted by Carstens and Dewey (2010). The tree shown is an estimate of the species tree obtained with a maximum likelihood method (STEM; Kubatko et al. 2009), with branch lengths given in coalescence units (generations in multiples of the population size, *N*). The support values on the branches, obtained with a Bayesian species tree method (BEST; Liu and Pearl 2007), show the posterior probability that the true species tree contains each of the clades shown (given the assumptions of the method).

The preceding methods are powerful and useful, but are constrained by the assumption that genealogical discordance is entirely due to incomplete lineage sorting. We know that other processes can cause discordance, most notably reticulation of population lineages (Chapter 6). How should we proceed when we are not willing to assume that the population history is entirely treelike?

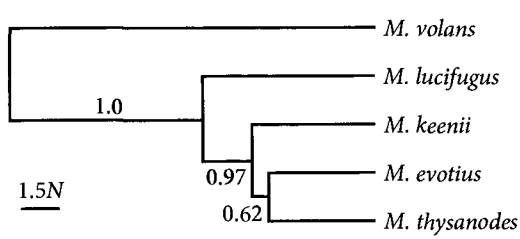


FIGURE 9.20 **Estimated species tree for five species of North American *Myotis* bats.** The inferred species tree topology is shown with branches drawn proportional to the branch's length in coalescent units. Numbers on internal branches are Bayesian posterior probabilities. Based on Carstens and Dewey (2010).

An alternative approach is to use a method that estimates the extent of genealogical discordance without committing to any one source of discordance. One such method is Bayesian concordance analysis, or BCA, which uses a simple measure of the prior probability of gene-to-gene discordance to convert sequence data from multiple genes into an estimate of the proportion of the genome for which any clade is true, its ***concordance factor***. Figure 9.21 provides a simple example: a concordance analysis of 30,000 gene alignments from humans and relatives (Ané 2010). This tree shows that for 76% (75–77%) of the genome, humans form a clade with chimpanzees, and that for over 99% of the genome, gorillas, chimpanzees, and humans form a clade. The 24% of the genome for which humans do not form a clade with chimpanzees is split evenly between the two other resolutions (see Figure 6.8), a pattern that is consistent with expectations under incomplete lineage sorting. This analysis finds no evidence of a process other than incomplete lineage sorting within the human-chimpanzee-gorilla clade.

A contrary result is summarized in Figure 9.22. A phylogenetic analysis of wild relatives of tomatoes based on 18 independent genes (Rodriguez et al. 2009) found three well-supported clades (core tomato, section *Juglandifolium*, and section *Lycopersicoides*). BCA suggested that of the three possible resolutions of these three clades, two are present at equal frequency among the sampled genes whereas one is completely absent. Statistical analysis of these results (Ané 2010) showed that this could not be explained by incomplete lineage sorting, suggesting instead that the core tomato clade likely derived from ancient hybridization between a member of each of the other two clades.

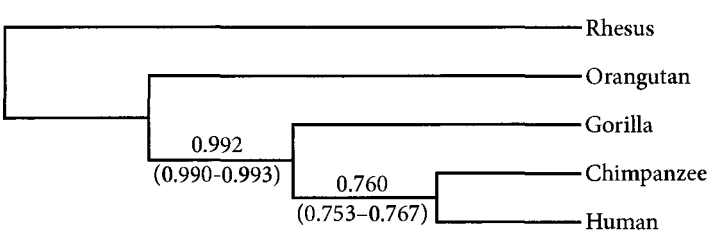


FIGURE 9.21 **Bayesian concordance analysis of 30,000 genes from five simian genomes.** The numbers above the branches are estimates of the concordance factor, the proportion of the genome that has that clade. The numbers below the branch are 95% credibility intervals on the concordance factors. As shown in Figure 6.8, of the ~24% of the genes for which chimpanzees and humans do not form a clade, ~12% show a human-gorilla clade and ~12% show a gorilla-chimpanzee clade. Adapted from Ané (2010).

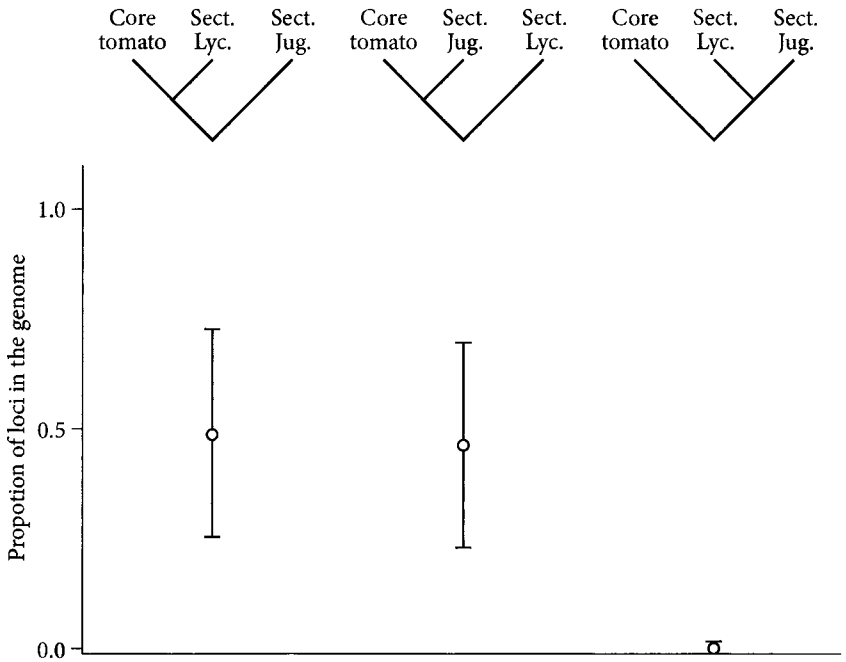


FIGURE 9.22 Bayesian concordance analysis of 18 genes from the tomatoes and their wild relatives. The three possible resolutions of the three well-supported clades (core tomato, section *Juglandifolium*, and section *Lycopersicoides*) are shown above. The graph below shows each clade's concordance factor and its 95% credibility interval. The fact that two of the possible resolutions have similar concordance factors, which are both significantly higher than the third resolution, is not predicted under incomplete lineage sorting. Instead, this result shows that the tomato clade could be of hybrid origin. Adapted from Ané (2010).

While BCA has the ability to detect cases in which reticulate evolution has occurred, it is best viewed as a stopgap solution. Model-based methods that incorporate processes other than incomplete lineage sorting should provide more accurate and powerful insights into evolutionary history. However, although significant progress has been made in modeling introgression between pairs of closely related species (Choi and Hey 2011), these models are too computationally demanding to allow for analysis of typical phylogenetic data sets. Nonetheless, in the not-too-distant future it may well become possible to use sequence data from many individual genes (or whole genomes) to

obtain detailed insights into the history of population reticulation. Such methods would represent an exciting advance, giving us new insights into the diversity of processes that contribute to genealogical discordance.

FURTHER READING

Tests of phylogenetic signal: Archie 1989; Faith and Cranston 1991; Hillis and Huelsenbeck 1992; Wilkinson et al. 2002

Measuring clade support: Templeton 1983; Kishino and Hasegawa 1989; Donoghue et al. 1992; Bremer 1994; Shimodaira and Hasegawa 1999; Goldman et al. 2000; Shimodaira 2002

Nonparametric bootstrap analysis: Felsenstein 1985a; Hillis and Bull 1993; Sanderson 1995

Parametric bootstrapping: Swofford et al. 1996; Huelsenbeck 1997; Huelsenbeck and Rannala 1997; Sullivan 2005

Bayesian statistical analysis: Nylander et al. 2004; Brown and Lemmon 2007

Comparing data sets: Farris et al. 1994; Huelsenbeck and Bull 1996; Sanderson and Shaffer 2002

Species tree methods: Maddison and Knowles 2006; Ané et al. 2007; Edwards et al. 2007; Knowles and Carstens 2007; Liu and Pearl 2007; Kubatko et al. 2009; Heled and Drummond 2010

CHAPTER 9 QUIZ

1. Suppose that you have analyzed a data set with maximum parsimony and have found several most parsimonious trees. What is the most important reason to do statistical analyses?
   a. To check whether there are more parsimonious trees that you did not find.
   b. To determine which of the most parsimonious trees is best.
   c. To see if the assumptions of parsimony apply for this data set.
   d. To help decide how confident you can be that features found in the most parsimonious trees are also found on the true tree.
   e. To prove that the most parsimonious trees are true and other possible trees are false.

2. Three of these four data sets are derived from evolution up a tree, whereas one is not. Which one is composed of random data?

```
   1                   10                  20                  30                  40
A  GGGACAGCAGAACGGTGATAGTCCTTGCAACGGCTGAATC
B  TTGACATCGGAGTGGCCGTAGCCCCTGTCATGCTTGCATC
C  GTGGCGTTGGCATGGTCACGTGCGTCGTCGTAGCTCCGGC
D  TTCAAAGTGAAACGCTGATAGTCGTTACAGTAGCCGCGTC
E  TGCAAATCAACACTCTGACGGTTGCTGTAACTCCCCCCGTT
F  GGTAAATTGACGCGCCCACATTCCTCGCAATCCCCCCAATC
G  GTCAAGTTGACACTGTGGCAGCTGTCGCAATTCCCCGCAGC
H  GTGGAATTGACACGGTGACAGTCGTTACAGTGCTCGAATT
```

a

```
   1                   10                  20                  30                  40
A  GCCGCGTTTCTTGTAATTGCCCAGTTGTTTATTAGAAATC
B  GCCGCGTTTCTTGTAATTGCCCAGTTGTTTATTAGAAATC
C  GTCACAGTCCTTGTAATCCCCCAGTCCTTAATTTCAAATC
D  GTGATAGGCGTCGCAATGCTCGCAGCCCTAGTGTCCACCT
E  GTGACAGTCGTCGCAATGCTCGCAGCCCTAGTTTCCACCT
F  CTGACAGTCGTTGCAGTGCCTGCATCCCCCAGGGTCACCTT
G  CTGATAGCCGCTACCGCACCTGCATCCCCCAGGGTCACCTT
H  CTGATAGCCGCCACCGCACCTGCATCCCCCAGGGTCACCTT
```

b

```
   1                   10                  20                  30                  40
A  TCTTGGTACGTAGCCCAAACCGACTGTGGACATCGGAACG
B  TCTCGGTACGTAGCCCAAACCGACTGTGGACGTCGGAACG
C  TCTCGCAAAATAGACCAACCCGACTGTGGACGTCGGAACG
D  GATAGCAAAATCAATCAACACGACGGGGTGGAAGTGACGTG
E  GATAGCAAAATCAATCAACACGACGGGGTCGAAGTGACGTG
F  GAAAACAAAACCAATCAACATACGGGGTCAAATTGACACG
G  GAAAACATAACCAATTGGCATACCGAGTCAAATTAACACT
H  GAAAACATAACCAATTGGCATACGGAGTTAAATTAACACT
```

c

```
   1                   10                  20                  30                  40
A  GTCACAGTCCTTGTAATCGCCCAGTCCTTAATTTCAAATC
B  CTGACAGTCGTTGCAGTGCCTGCATCCCCAGGGTCACCTT
C  GCCGCGTTTCTTGTAATTGCCCAGTTGTTTATTAGAAATC
D  GTGACAGTCGTCGCAATGCTCGCAGCCCTAGTTTCCACCT
E  GTGATAGGCGTCGCAATGCTCGCAGCCCTAGTGTCCACCT
F  CTGATAGCCGCCACCGCACCTGCATCCCCCAGGGTCACCTT
G  CTGATAGCCGCTACCGCACCTGCATCCCCCAGGGTCACCTT
H  GCCGCGTTTCTTGTAATTGCCCAGTTGTTTATTAGAAATC
```

d

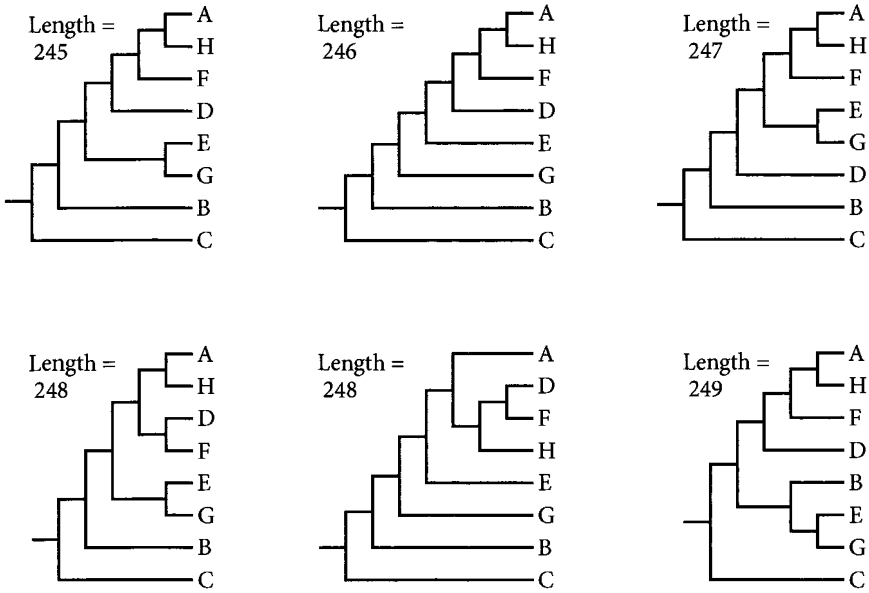
3. Why should a data set that evolved up a tree tend to yield a shorter most parsimonious tree than a data set that has similar characteristics (e.g., number of variable and parsimony-informative characters) but where the tips were generated independently without evolution up a tree?
   a. Because natural selection tends to remove traits that lengthen trees.
   b. Because evolution along a tree reduces the number of variable characters.
   c. Because only the data that evolved up a tree will yield a most parsimonious tree; the random data will yield unparsimonious trees.
   d. Because evolution up a tree will tend to yield characters that are consistent with one another.
   e. The premise is false; we do not expect a difference in the length of the most parsimonious trees from the two data sets.

4. Gene sequences for eight species are found to have a most parsimonious tree of length 460. A PTP test is conducted and yields the distribution seen in the table below. Which of the following best explains this result?
   a. The species share descent from common ancestry.
   b. The species evolved along a tree, but there was an error in the laboratory such that the genes are not correctly associated with the species.
   c. The species evolved along a tree, but by chance there was less homoplasy than expected.
   d. The species evolved along a tree, but the genes evolved so rapidly that the evolutionary signal has been obscured.
   e. The genes have tracked the species tree and have evolved in such a way that significant treelike structure has been retained.

Length	Frequency	Length	Frequency
443	1	456	133
444	1	457	110
447	5	458	117
448	5	459	81
449	17	460	64
450	21	461	61
451	26	462	28
452	42	463	23
453	72	464	9
454	76	465	3
455	101	466	2

5. What is a decay index?
   a. The number of PTP replicates that have a given clade.
   b. A measure of how strongly the data support a given clade divided by the prior probability of that clade.
   c. A measure of the probability that a clade is false.
   d. The number of bootstrap replicates in which the clade has disappeared.
   e. The length of the shortest tree that lacks a clade minus the length of the most parsimonious tree.

6. The figure shows the six shortest trees for a given data set. What clade has a decay index of 3?
   a. (F, D) b. (E, G) c. (A, H, F, B) d. (A, H, F) e. (D, F, A, H)



7. You do an unconstrained parsimony search and find that the shortest trees all have length 3275 and all have clade X. You then do a constrained parsimony search and find that the shortest tree that lacks X has length 3274. How do you interpret this result?
   a. The decay index of X is 3274.
   b. The decay index of X is 3275.

c. The decay index of X is 1.
d. The constrained search failed to find the most parsimonious trees.
e. The unconstrained search failed to find the most parsimonious trees.

8. What does a bootstrap score of 70% for a particular clade mean?
   a. 70% of the bootstrap replicates yield a most parsimonious tree with that clade.
   b. 70% of the genes support the existence of this clade.
   c. 70% of the equally most parsimonious trees have that clade.
   d. 70% of the characters in the data set are consistent with that clade (30% disagree with that clade).
   e. There is a 70% chance that the clade is on the true tree.

9. How are nonparametric bootstrap replicates generated?
   a. By randomly evolving characters up a tree.
   b. By randomly reordering the characters in a matrix.
   c. By randomly sampling characters, with replacement, from the matrix.
   d. By randomly subsampling the data matrix so that half of the characters are not sampled at all.
   e. By randomly selecting characters while ensuring that no character is picked more than once.

10. Here is a small data matrix. Create a possible permuted data set such as might be used for a PTP test.

Taxon	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
A	A	T	G	T	C	T	T	G	C	T	A	C	A	G	A
B	A	T	A	T	C	T	T	A	C	T	G	C	G	C	A
C	A	T	G	C	C	A	A	A	C	T	A	C	G	C	A
D	A	G	G	C	T	A	A	A	A	T	G	T	A	T	G
E	A	G	A	C	C	A	G	G	A	T	G	T	A	T	A

11. Using the data matrix in Question 10, create one possible bootstrap replicate.

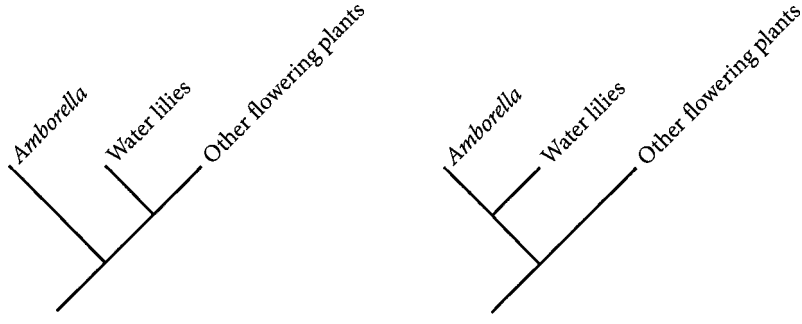
12. You have sequenced three different genes for the same set of taxa. Parsimony searches for each gene separately yield most parsimonious trees of length 72, 322, and 251 (sum = 645). You do a partition homogeneity test with 10 replicates, generating 10 random assignments of characters to the three partitions (of the same size as the original genes). The length of the three partitions for the randomized data partitions is shown in the table.

Replicate number	Length of partition A	Length of partition B	Length of partition C	Sum of lengths
1	78	301	258	637
2	80	333	235	648
3	71	329	254	654
4	68	312	263	643
5	73	352	247	672
6	69	307	271	647
7	77	329	245	651
8	72	341	229	642
9	79	310	244	633
10	66	318	258	642

a. Should anything be read into the fact that partition A is always associated with a shorter tree than the other partitions?

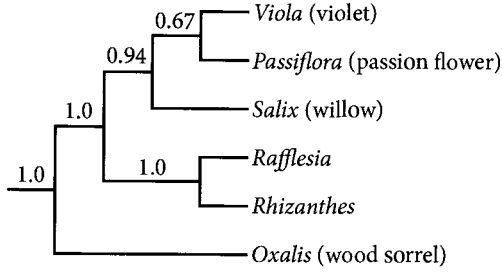
b. Does this small sample of random partitions tend to support or reject the hypothesis that the three genes have different evolutionary histories?

13. Zanis et al. (2002) investigated the relationships at the base of the flowering plant clade. Most prior studies supported tree 1 with a New Caledonian plant, *Amborella*, as sister to the rest of the flowering plants. A few other studies (e.g., Barkman et al. 2000) supported tree 2, in which *Amborella* + water lilies are the sister group to other flowering plants. Zanis et al. simulated new molecular data sets along tree 2 and found that 84% of these data sets returned tree 1 in parsimony analyses. What term best describes this test? What does the result suggest?



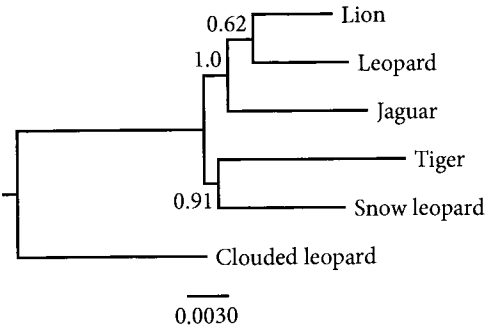
Tree 1 Tree 2

14. Inferring phylogenies for parasitic plants poses a challenge for systematists because parasitic plants usually undergo drastic changes in their genomes and in their morphology, making them hard to compare with other taxa. Barkman et al. (2004) used mitochondrial DNA to place *Rafflesia* (the genus with the world's largest flower) in the flowering plant phylogeny. Posterior probabilities are shown above branches. In a Bayesian framework, does this analysis reject the possibility that *Rafflesia* is more closely related to *Oxalis* than to *Viola*?



15. Explain why minimizing deep coalescence makes sense as a criterion for identifying the best species tree given a set of gene trees. Why should a species tree with more deep coalescence events be a worse estimate of the population history?

16. The figure is an estimate of the species tree for some large cats (based on Davis et al. 2010). The numbers on the internal branches are the posterior probability that the clade exists on the species tree. Assume that the internal branch lengths are proportional to time and that population size has remained more or less constant. Would you expect a higher proportion of the genome to support a sister group relationship between lion and leopard or between tiger and snow leopard? Why?



CHAPTER TEN

Using Trees to Study Character Evolution

Phylogenetic trees have uses beyond structuring classifications (Chapter 5). They also provide powerful tools for learning about evolutionary history. A phylogeny may seem to contain limited information, namely, the order of lineage-splitting events and, depending on the inference method, the relative time between those events. However, we can use information on tree topology and branch lengths to learn about where and when different organismic features evolved. By combining these inferences with scientific models of how one trait affects and is affected by the evolution of other traits, it becomes possible to learn something about *why* evolution unfolded the way it did.

In this chapter, we start by describing how phylogenetic trees can be used to test adaptive hypotheses. Then we introduce model-based statistical methods for reconstructing the evolution of discrete traits and show how they can be used to detect correlated evolution of suites of traits. Then we turn our attention to continuous traits, first introducing statistical models for their evolution and finally showing how these models can be used to assess whether two continuous traits have evolved independently.

HISTORICAL CHRONICLES AND NARRATIVES

Philosophers of history draw a distinction between a historical chronicle, a listing of what events occurred and when, and a historical narrative, an analysis of why those events happened as they did. A chronicle of the Second World War might include a listing of battles, the dates and content of the speeches of Adolf Hitler, Winston Churchill, Franklin Roosevelt, and other political leaders, information on weather and food production for each month, and so on.

This contrasts with a narrative, which provides an explanation for *why* certain events happened. For example, a narrative might explain why Britain declared war on Germany in 1939, why Russia signed a nonaggression pact with Germany, and why Germany was ultimately defeated. By applying general models of human behavior, economic principles, and so forth, we can arrive at well-corroborated historical narratives. But our ability to develop compelling historical narratives depends on starting with an accurate historical chronicle.

As O'Hara (1988) so eloquently noted, the same structure can be applied to evolutionary biology. When building historical narratives in evolutionary biology, for example, explaining why a certain trait evolved, we depend on an accurate chronicle, a good estimate of the true phylogenetic history.

> Anyone who has internalized tree thinking will recognize that attempting to answer a question like "Why are flamingos pink?" without significant knowledge of the evolutionary chronicle of flamingos (and probably a number of other clades of birds as well) is as foolish as attempting to explain why England declared war on Germany without knowing whether Germany invaded Poland or Poland invaded Germany.

As described in Chapter 4, the principle of parsimony can guide us in making inferences as to when a particular trait evolved, which amounts to reconstructing an evolutionary chronicle. How can we use this chronicle to assist us in building an evolutionary narrative, an explanation for why certain traits evolved?

When an evolutionary biologist asks *why* a certain trait evolved, the question boils down to the role of natural selection in trait origin. Was the trait's origin shaped by selection and, if so, for what function? A trait that evolved by natural selection for some biological function is considered an ***adaptation*** for that function. Thus, testing a hypothesis of adaptation amounts to testing a historical narrative. As outlined in the next section, the study of adaptation is a rigorous discipline that depends heavily on a phylogenetic perspective.

TESTING HYPOTHESES OF ADAPTATION

The hypothesis that a derived trait is an adaptation is the claim that the trait evolved because selection favored ancestral organisms that had the derived trait. Testing this hypothesis requires evaluating why a population lineage that was polymorphic (Chapter 3) for ancestral and derived versions of the trait became fixed for the derived character state and lost the ancestral state. Before

exploring tests of adaptive hypotheses, it is worth noting that some hypotheses are invalid and not even worth trying to test.

First, a valid adaptive hypothesis should not confound cause and effect. It is no more reasonable to say that the Allies won the Second World War *because* this outcome allowed the division of Europe into an Eastern and Western bloc than to say that body hair evolved in mammals *because* porcupines could later modify their hairs into protective spines. Evolution does not plan ahead: valid adaptive hypotheses should consider only the effect of traits at the time they originate.

Second, an evolutionary narrative is only useful if it is specific enough to be informative. To say that the Allies won because they had a stronger army, where "stronger" just means more prone to winning, is not testable, because it really cannot be wrong. Likewise, to say that body hair evolved in mammals because it was "selected" is too permissive. Rather, a good adaptive hypothesis specifies a biological function that was enhanced by the trait in question so as to confer a selective advantage. For example, we might propose that mammalian body hair is an adaptation for thermoregulation. Or, to be more explicit, that alleles that resulted in the production and increased density of body hair in early mammals were favored (over alleles conferring less hair) because individuals with these alleles had better body temperature regulation.

Once we have a valid adaptive hypothesis, how can we test it? Furthermore, how can we compare one hypothesis to an alternate one? For example, how could we evaluate whether it is more likely that body hair evolved because of thermoregulation or because it afforded early mammals protection from sunlight?

To provide a concrete basis for the presentation, let us consider two derived traits of humans: opposable thumbs and mobile shoulder joints. The adaptive hypothesis we will test for the thumbs is that they evolved because organisms with opposable thumbs were better able to manipulate tools. Similarly, we will test the hypothesis that a mobile shoulder evolved because it improved throwing ability.

Before proceeding further, we need to clarify the nature of the traits in question. The main anatomical feature that allows humans to make thumbtip to fingertip contact with any finger is that we have a saddle joint connecting the first thumb bone to the wrist bone. This joint is situated in your palm near the wrist. Other primates have a different kind of joint, which provides for greater hand strength but less mobility. Likewise, several features of the human shoulder

allow for its amazing flexibility, but the placement of the shoulder blade (scapula) is one of the most important. Whereas many primates have the scapula situated laterally (on the side of the body), we have a scapula that is situated dorsally (closer to the center of the back). A dorsal scapula allows our arms to have a wide range of movement.

The first prediction an adaptive hypothesis makes is that the trait is beneficial for the function in question: the derived character state is better than the ancestral character state at performing the biological role. Once we have confidently identified the ancestral character state (using a tree and parsimony, or some other method), the key test is functional. We can use experimentation or theoretical modeling to assess whether an organism with a saddle joint and an opposable thumb is better able to manipulate tools than one without such a joint. Similarly, we would need to assess whether a primate with dorsal scapulae is better at throwing than one with lateral scapulae. Let us assume that, in both cases, the functional analyses come back in the affirmative—both traits enhance performance of the specified biological function.

The second prediction of an adaptive hypothesis is that the trait evolved at a time when improving performance of the biological function conferred a fitness advantage. It does not matter how much better it is to have an opposable thumb for typing on a computer keyboard, we *know* that opposable thumbs are not adaptations for that role because they evolved well before an ability to type on a keyboard conferred a selective advantage (if it even does now!). The set of ecological factors that determine whether a trait variant is favored by selection is called the ***selective regime.*** Can we reconstruct the selective regimes of ancestors?

Ancestral state reconstruction allows one to make inferences about any biological attribute that has some degree of heritability. Selective regimes will often be heritable because the action of selection on one trait is shaped to a large extent by the suite of other traits that a lineage already has. As a result, it is often possible to use parsimony or likelihood methods to infer the selective regime that was in effect when a trait arose.

Figure 10.1 shows the phylogeny of the simian clade of primates. Over this we have provided the parsimony mapping of opposable thumbs and tool use for food collection. The latter is the selective regime under which improved tool manipulation would be expected to confer a selective benefit. While this analysis was conducted using parsimony, the rarity of the derived traits allows us to be quite confident in this reconstruction.

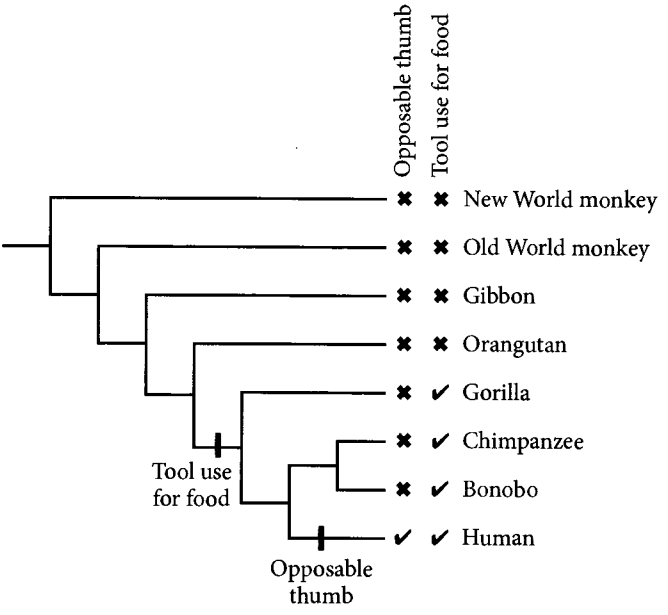


FIGURE 10.1 Phylogeny of the simian primates with two traits, opposable thumbs and tool use, mapped onto the tree using equally weighted parsimony. The inference that an opposable thumb evolved after tool use is consistent with the hypothesis that an opposable thumb is an adaptation for more effective use of tools.

Based on Figure 10.1, we can conclude that the opposable thumb arose in a lineage that was already using tools to assist in food gathering. We know this because chimpanzees and gorillas frequently use tools, for example, using a stick to extract termites from a termite mound. This pattern suggests that when the opposable thumb evolved on the terminal lineage leading to living humans, the population was already using tools for functions that influenced fitness. Thus, if we assume that an opposable thumb improves one's ability to use tools, it is reasonable to suppose that selection for improved tool use contributed to the evolution of opposable thumbs. The finding that the opposable thumb evolved after tool use directly supports the claim that the opposable thumb evolved *because* it aided in tool use.

Figure 10.2 shows parsimony mapping of the dorsal scapula and the use of projectiles for hunting, the selective regime under which improved throwing

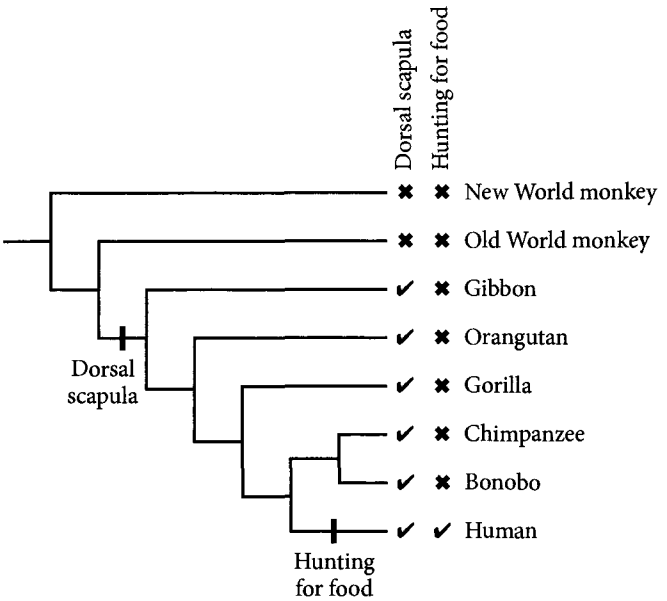


FIGURE 10.2 Phylogeny of the simian primates with two traits, dorsal scapula and hunting, mapped onto the tree using equally weighted parsimony. The inference that dorsal scapulae evolved before projectile use for hunting shows that dorsal scapulae are not an adaptation for throwing projectiles. Instead, the dorsal scapula trait can be considered an exaptation for throwing.

would be expected to confer a selective benefit. This figure shows that the shoulder trait evolved earlier than the origin of active hunting. This is supported by the fact that similar shoulder modifications are found in all apes, whereas only humans commonly hunt with projectiles (chimpanzees have been observed using rocks to kill small animals, but they do not hunt with projectiles). By showing that hunting with projectiles evolved well after the dorsal scapula, we have effectively refuted the claim that the dorsal scapula evolved because it improved hunting—the adaptive hypothesis is therefore refuted by the phylogenetic data.

Instead of being an adaptation for throwing, dorsal scapulae and mobile shoulders appear to have evolved for another function, such as hanging from and swinging beneath branches. While the trait became beneficial for throwing projectiles, this was a fortuitous feature of an already existing trait rather than

an explanation for the trait's origin. Traits that are now useful for a function for which they did not originally evolve are called ***exaptations.***

The examples of opposable thumbs and mobile shoulder joints in humans serve to illustrate the general principle that trees and tree thinking can be combined with functional and ecological studies to help us learn why specific traits originally arose. Through the use of phylogenies we can ask "why" questions, without which evolutionary biology would be a much less interesting subject.

THE COMPARATIVE METHOD

In the preceding section we discussed traits, such as opposable thumbs, that evolved in one specific place on the tree of life. It is often the case, however, that functionally similar traits evolve repeatedly across the tree. Although these traits are not homologous, they might all have arisen via similar adaptive causes. This gives us an especially powerful approach to studying adaptation, often called the ***comparative method.*** By looking across multiple species, we can determine if there is a statistically significant correlation between the possession of a certain trait and occupation of a particular selective regime (or a second trait that may be indicative of a selective regime). The comparative method is also used to test for correlations between continuous traits, but discussion of this application will be postponed until the end of the chapter.

A significant correlation detected in a comparative analysis may indicate adaptation, showing that whenever a lineage shifted into a certain selective regime, natural selection "drove" the origin of a similar trait. However, it should be noted that there are other possible explanations. If two traits are developmentally interrelated, then whenever one trait evolves the other one may arise as a developmental by-product. For example, finding that the character states *fur pale* and *melanin levels low* are correlated in a survey of mammal species might not tell us about adaptation per se. It may merely confirm that melanin is the main pigment in mammal hair. That being said, evidence of a correlation between a pair of traits can provide powerful evidence of selection.

In the early days, the comparative method was often employed without reference to phylogeny. A count of species with different combinations of traits was made and a standard statistical test was used to judge significance. However, Felsenstein (1985b) pointed out that the null hypothesis of such statistical tests is complete independence. But so long as the two traits have evolved along

the same phylogenetic tree, they are not strictly independent. This, he showed, can lead to spurious results.

Suppose we look at the habitat (forest versus prairie) and fur tone (dark versus pale) in a hypothetical clade of mammals. We count up species (tips) with each trait and find the result shown in Table 10.1. Applying a chi-square test to this contingency table yields a very significant result. So you might conclude that dark fur is an adaptation to dim forest interiors. However, the significance of this ***tip correlation*** does not mean that there is a meaningful ***evolutionary correlation*** once phylogeny is taken into account.

Figure 10.3 illustrates a phylogenetic history that would undermine the conclusion that the traits in Table 10.1 have an evolutionary correlation. This tree shows that the apparent deviation from the null expectation is attributable to a pair of large clades: one dark-haired clade living in forests and one light-haired clade living in prairies. Since four other lineages appear to have persisted as long as the large clades and to have the other combinations of traits (forest-pale and prairie-dark), it seems that there is no significant evolutionary correlation between habitat and fur color. While this tree does not necessarily disprove the claim that dark fur is an adaptation to a forest habitat, it shows that the observed tip correlation does not provide evidence for an evolutionary association between these two traits.

There are a number of statistical methods for assessing whether the association between traits and selective regimes, or between pairs of traits, is significant after taking phylogeny into account. These may be conducted in the context of the best estimate of a group's phylogeny. Alternatively, if there is uncertainty in the tree topology or branch lengths (as is almost always the case), the same comparative method can be run on a sample of plausible trees, for example, trees obtained from bootstrap analysis or a Bayesian posterior sample obtained from MCMC.

To be as rigorous as possible, comparative methods should also account for uncertainty in ancestral state reconstruction. To provide a flavor for the think-

TABLE 10.1 Contingency table showing the distribution of habitat and fur traits in a hypothetical group of mammals.

	Forest	Prairie
Pale fur	2	100
Dark fur	150	2

ing, we will here introduce a rather easy-to-understand method: the ***concentrated changes test.*** Because this method uses parsimony, it does not incorporate uncertainty in ancestral states.

A typical implementation of the concentrated changes test maps the evolution of a selective regime onto a tree and then counts up the number of times that a trait of interest arose on branches that occupied the ancestral or the derived selective regimes. Applied to Figure 10.3, parsimony suggests two origins of dark fur, one in a prairie habitat and one in a forest habitat. This pattern does not support the hypothesis that there is a bias toward more origins of dark fur in forests (or more origins of pale fur in prairies). Thus, the concentrated changes test would not support an evolutionary correlation between fur color and habitat in this case.

As a more realistic example, consider a study of a large group of spiders (Miller 2007). This research was designed to ask whether there is significant correlation between male sacrifice behavior, wherein males die or are killed by their mate at copulation, and males having genitals that typically become broken or otherwise mutilated the first time they are used for copulation. Evolutionary

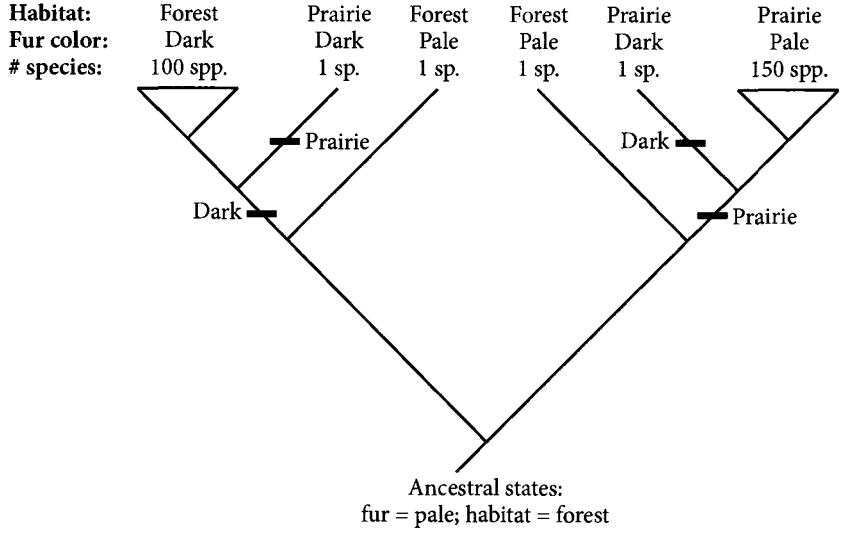
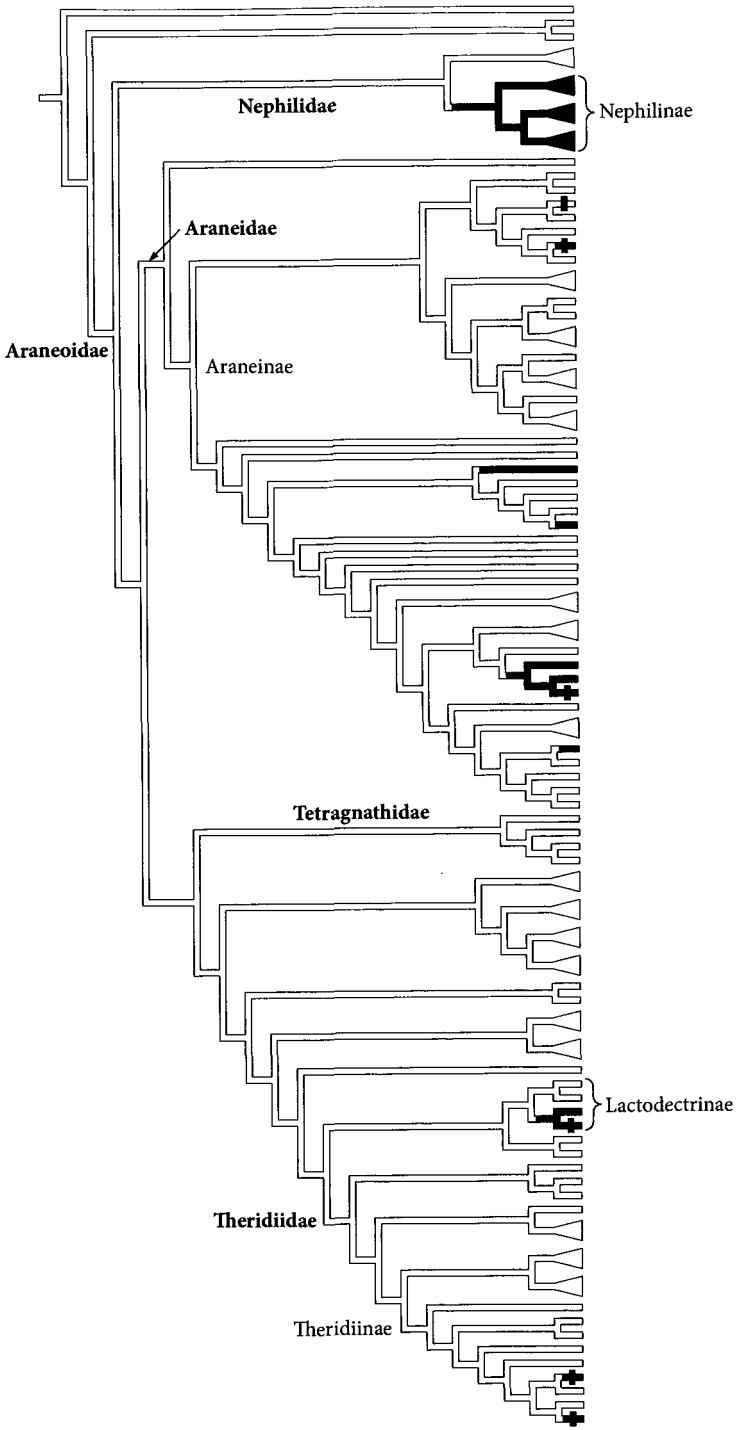


FIGURE 10.3 A hypothetical pattern that could lead to a misleading significant result in a chi-square test. The maximum parsimony mapping of the two traits is shown. While there has not been correlated evolution of habitat and fur color, the presence of two large clades that lack variation in either trait yields a significant chi-square test when species are used as tips.


Nephilidae
Nephilinae
Araneidae
Araneoidae
Araneinae
Tetragnathidae
Lactodectrinae
Theridiidae
Theridiinae

theory suggests that the presence of genital mutilation should favor the evolution of male sacrifice: if a male is doomed to mate only once, he can increase his reproductive output by sacrificing himself fully to the female, providing her sustenance and thus increasing the potential success of his offspring.

Figure 10.4 shows a phylogeny for spiders with the two traits mapped by parsimony. Of the six origins of male sacrifice, five occur in branches that are mapped as having male genital mutilation. Is this a meaningful correlation between these two traits, or could such a pattern arise by chance? Given that male sacrifice evolved six times on this tree, what is the probability that five (or more) of those events would have landed on branches that had genital mutilation? We can answer this by considering a large number of random simulations of the evolution of male sacrifice on this tree, all entailing six origins of the trait, to see how often five or more of those origins land on branches that are mapped as having genital mutilation. With the help of the computer program MacClade (Maddison and Maddison 2000), we can show that the probability of such a concentration of changes by chance is very low, about 0.0008. This convincing phylogenetic correlation is consistent with the theoretical predictions, and helps to build a compelling evolutionary narrative that at least partly explains why certain male spiders become their mate's dinner.

Although the concentrated changes test is intuitive and easy to implement, it relies on parsimony reconstructions, which have three undesirable features for comparative studies. First, parsimony does not take account of the duration of branches. This is unrealistic because, all things being equal, an evolutionary transition is more likely to occur on a branch of long rather than short duration. Second, parsimony does not allow us to say whether the cost of gaining a character state should be equal to the cost of losing it. Or, more generally, there is no way to use the data to select an appropriate step matrix for generalized parsimony (see Chapter 7). Third, parsimony does not provide an easy measure of statistical confidence in the assignment of states to ancestral nodes. For a character with two character states, parsimony provides only three possible conclusions about a node: it was state 0, it was state 1, or it is equivocal (i.e.,

FIGURE 10.4 Spider phylogeny with male genital mutilation and male sacrifice mapped onto the tree by parsimony. (*Opposite*) Black branches have male genital mutilation, whereas horizontal black bars indicate origins of male sacrifice behavior. Tips composed of multiple taxa are shown with triangles. There are six independent origins of male sacrifice and nine independent origins of male genital mutilation. Five of the six origins of male sacrifice occur on lineages that are inferred to have genital mutilation, despite the fact that relatively few branches are mapped as having genital mutilation. This is a significant association based on a concentrated changes test. Modified from Miller (2007).

some equally parsimonious reconstructions assign state 0 to the node whereas others assign state 1). It is not possible to use parsimony to say, for example: "this node has a 70% chance of having been state 0 and a 30% chance of having been state 1."

Parsimony-based approaches, such as the concentrated changes test, are becoming less common in comparative studies. Their principle use is in cases where we lack a phylogeny with branch lengths, a situation that can arise when it is necessary to piece together a phylogeny from previous studies based on different data sources. However, even in such situations, practitioners have developed ways to generate approximate branch lengths in order to allow for the use of model-based approaches.

MODEL-BASED ANALYSIS OF THE EVOLUTION OF DISCRETE TRAITS

Before we can describe the model-based analogs of the concentrated changes test, we first need to introduce probabilistic models for how discrete traits evolve along the branches of a known phylogeny. These models are similar to the models used to describe trait evolution during maximum likelihood and Bayesian inference of phylogenies (Chapter 8).

Let us start by considering a simple, binary character such as the presence or absence of wings, denoted 0 and 1, respectively. This can be modeled using a two-by-two rate matrix, as shown in Table 10.2. Instead of μ (the rate of DNA substitution), the rate of character evolution of a binary character is traditionally denoted q. If we assume that the rate of gaining and losing wings is equal (we will relax this assumption later), the substitution (or "transition") probability matrix is as shown in Table 10.3. You will probably note the similarity to the Jukes-Cantor model, with two states instead of four. This model is also commonly portrayed in a pictorial form (Figure 10.5).

Having defined this simple model of trait evolution, it becomes possible to use maximum likelihood to make inferences about trait evolution along a prespecified tree with branch lengths. The analysis aims to find the rate of character evolution, q (scaled to branch length), that maximizes the likelihood (the probability of the observed trait data). The value of q that yields the highest likelihood will depend on the tree and the distribution of states among taxa.

TABLE 10.2 Symmetric rate matrix for a binary character

From	To	
	0	1
0	$-q$	q
1	q	$-q$

TABLE 10.3 Substitution probability matrix for a binary character

From	To	
	0	1
0	$\frac{1}{2} + \frac{1}{2}e^{-qt}$	$\frac{1}{2} - \frac{1}{2}e^{-qt}$
1	$\frac{1}{2} - \frac{1}{2}e^{-qt}$	$\frac{1}{2} + \frac{1}{2}e^{-qt}$

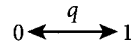


FIGURE 10.5 Model of evolution for a single binary character. The parameter q describes the rate of transition between the two states of the character (0 and 1).

As with maximum likelihood analysis of sequence evolution, the calculation of the probability of the trait data involves summing over all of the possible ancestral states at all nodes. For example, for a rooted tree with 20 taxa, every combination of 0s and 1s needs to be considered for all 19 nodes. The total probability of the trait data (for a given value of q) is a summation over all 2^{19} distinct trait histories. Each of those trait histories breaks the tree down into a set of branches (37 for 20 taxa), each of which has a specified branch length and states assigned to the two ends. For a given rate of change, it is easy to use the substitution probabilities (Table 10.3) for each branch and combine them to calculate the probability of the tip data evolving. Then, by adjusting q, we can find the value that maximizes this likelihood.

Once we have found the maximum likelihood value of q, we can ask for each node on the tree, How much of the total likelihood is attributable to histories in which this node had state 0, and how much is attributable to trait 1? These relative (or conditional) likelihoods provide information as to how confident we should be that a node had state 0 or state 1.

A common way to show the relative likelihoods is with a pie graph placed at the node. As a rule of thumb, if the circle is more than about 7/8 filled with one of the colors, we can be confident at the approximately $P = 0.05$ level that this is the correct ancestral state (given the tree, model, branch lengths, and

trait data). To give you an intuition for these principles, consider two examples (both from Schluter et al. 1997).

The first case involves the evolution of gall induction by wasps. Gall wasps lay their eggs inside plant tissue and induce the plant to make a protective gall around the larva. Some genera of gall wasps, however, lack the ability to induce galls and instead lay their eggs within the gall formed by another wasp. These so-called inquiline genera have been inferred to fall into a single clade, as shown in Figure 10.6.

Because a single clade contains taxa with state 1 and the rest of the tree only contains taxa with state 0, maximum likelihood analysis on this tree favors a low rate of evolution. If the rate of change is sufficiently low, then it is most plausible that there was just a single change of character state, and that this state change was at the base of the inquiline (state = 1) clade. This can be seen when we look at the relative likelihood of the gall-forming and inquiline habit at each node in the tree. You can see that it seems almost certain that the inquiline habit evolved on the stem lineage of the clade that includes all the inquilines. Because the estimated rate of trait evolution is low, the maximum likelihood criterion yields a similar conclusion to parsimony.

A different kind of result arises when taxa with 0 and 1 are scattered across the tree. In such cases maximum likelihood favors a model in which the rate of

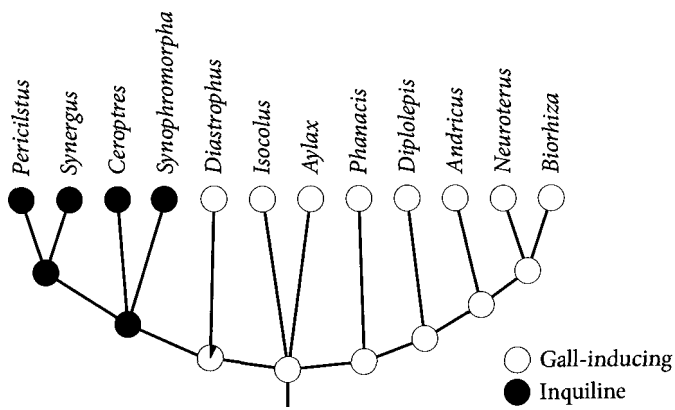


FIGURE 10.6 Ancestral state reconstruction of galling behavior in cynipid wasps based on maximum likelihood. Pie charts show the relative likelihood that an ancestor manifests the galling or inquiline habit.

character evolution is high. High rates of evolution imply that many changes of state occurred, only to be obscured by subsequent evolution. This tends to result in greater uncertainty in the assignment of states to nodes and greater deviation from parsimony.

An example involving a moderately high rate of trait evolution is provided by the fish genus *Xiphophorus* (platies and swordtails). In many species, males have an elongated tail, a "sword," which seems to play a role in sexual selection analogous to the tail of male peacocks. Based on a molecular phylogeny, there are no major clades that include just sword-bearing species. Rather, swords appear to have been gained and/or lost several times. As summarized in Figure 10.7, because a high rate of tail-shape evolution is inferred, many nodes of the tree cannot confidently be called as either having or lacking swords.

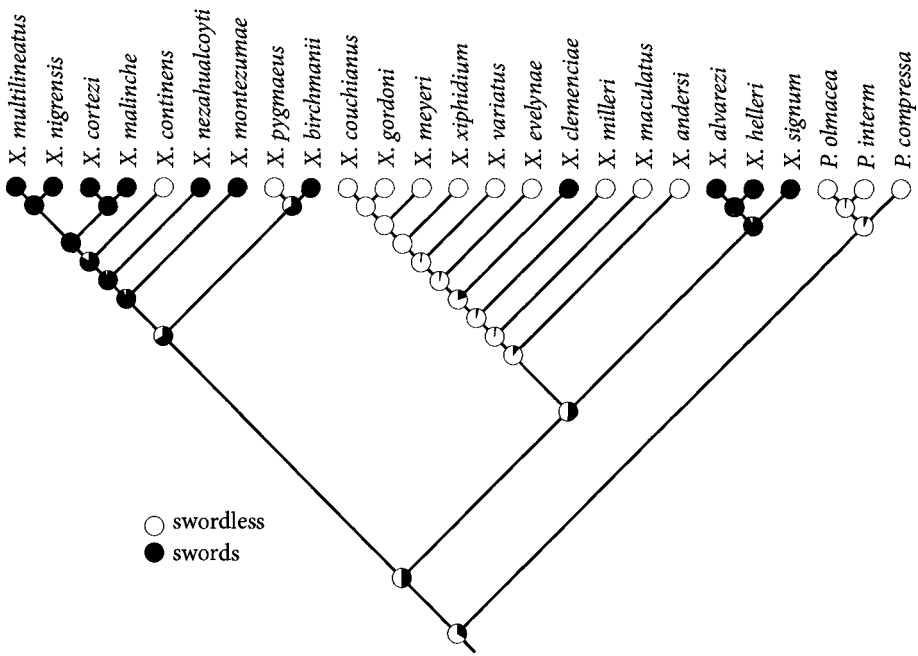


FIGURE 10.7 Ancestral state reconstruction of sword-tail presence/absence in the fish genus *Xiphophorus* based on maximum likelihood. Pie charts show the relative likelihood that an ancestor had or lacked a sword tail.

EVOLUTIONARY TRENDS

When using maximum likelihood, it is not necessary to assume that the rate of going from state 0 to state 1 (e.g., swordless to sword bearing) is the same as the rate of going from state 1 to state 0 (e.g., sword bearing to swordless). Sometimes evolution may have inherent directionality, a ***trend.*** Trends are important to be aware of because if we erroneously assume that there is not a trend when there is, we may arrive at an inaccurate evolutionary chronicle.

An evolutionary trend can arise for many reasons. One that is easy to imagine is a selective bias. Suppose that presence of the trait (state 1) is heavily favored in environment A and is neutral in environment B, whereas the absence of the trait (state 0) is slightly favored in environment B and is strongly disfavored in environment A. In this case, as lineages move back and forth between A and B, we expect more gains (changes from 0 to 1) than losses (changes from 1 to 0). Alternatively, there could be a mutational bias. If populations fixed for trait presence commonly experience mutations that result in trait loss, then we might expect to see many evolutionary losses and relatively few gains.

Whenever there is a gain-loss bias, we expect an evolutionary trend to be seen: a clade that starts with an equal number of tips in each state will tend to accumulate more of the state that arises at a higher rate. For example, if the rate of gaining a trait were three times the rate of losing it, the expectation is that at equilibrium ¾ of the lineages will have the trait and ¼ will lack it.

Using likelihood methods, it possible to consider asymmetric models in which the rate of gains and losses are allowed to be different. Instead of one rate of change q, we define separate rates of trait gain (q_{01}) and trait loss (q_{10}), as shown in Table 10.4. This is shown graphically in Figure 10.8.

In asymmetric models, the rate of gain and loss are allowed to differ, but they are not forced to differ. If the data happened to favor $q_{01} = q_{10}$, then the likelihood score of the asymmetric model would be identical to the symmetric model (Table 10.2). This shows that the likelihood of the asymmetric model will always be equal to or higher than the symmetric model. Based on a set of data, a likelihood ratio test (Chapter 8) can be used to see if the likelihood of the asymmetric model is so much higher that the symmetric model is implausible. If the data fail to justify using the more complex asymmetric model, statistical theory would advocate selecting the simpler symmetric model.

TABLE 10.4 Asymmetric rate matrix for a binary character

From	To	
	0	1
0	$-q_{01}$	q_{01}
1	q_{10}	$-q_{10}$

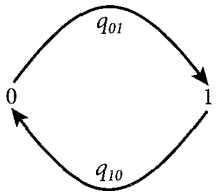


FIGURE 10.8 Evolutionary model for a single binary trait that allows for different rates of gain and loss. The parameter q_{01} describes the rate of gains (0 to 1 changes), whereas q_{10} describes the rate of losses (1 to 0 changes).

It is a strength of maximum likelihood relative to parsimony that the data can be used to assess whether there is a gain-loss bias. Furthermore, if the data support such a directional trend, maximum likelihood can be used to estimate the ratio of the two rates.

Figure 10.9 shows the results from a study of floral symmetry evolution, based on a phylogeny containing 379 species of flowering plants (Ree and Donoghue 1999). The authors used maximum likelihood methods to estimate the transition rate from radial symmetry (actinomorphy) to bilateral symmetry (zygomorphy) and vice versa. The horizontal axis shows the rate of gaining zygomorphy and the vertical axis shows the rate of losing zygomorphy (evolving actinomorphy). The shading represents the likelihood score given the rates shown, with lighter colors referring to higher likelihoods. The symmetric model, where the rates of gaining and losing zygomorphy are forced to be equal, corresponds to the diagonal line. Under the symmetric model, the highest likelihood is indicated with a square. The asymmetric model, which allows for rates to be drawn from anywhere on this surface, has its highest likelihood when there is a high rate of losses (evolving from zygomorphy to actinomorphy) and a low rate of gains. This combination is marked with a circle. This implies that zygomorphy is hard to gain but easy to lose. One possible explanation is that mutations converting zygomorphic flowers into actinomorphic flowers are relatively common, whereas the reverse mutations are relatively rare.

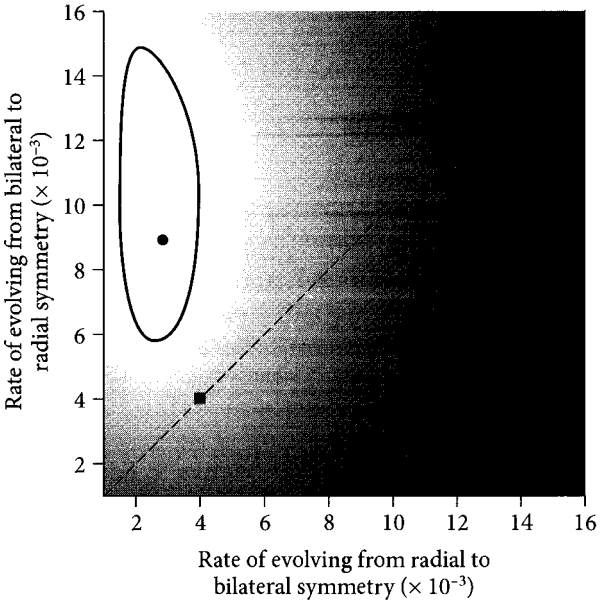


FIGURE 10.9 Probabilities of the data given different rates of gaining and losing bilateral symmetry (zygomorphy) in asterid flowers. The shading represents the likelihood score given the rates shown, with lighter colors referring to higher likelihoods. The small circle marks the maximum likelihood estimate of the two rates, whereas the contour line around this peak marks the 95% confidence interval around these rates. The diagonal line corresponds to the symmetric model in which q_{01} equals q_{10}, with the square marking the maximum likelihood rate estimate under this model. Adapted from Ree and Donoghue (1999).

Given the tree, branch lengths, and trait scoring, we can place a confidence interval around our estimates of the rates of evolution in each direction. This is shown with the contour line around the optimal value in Figure 10.9. Note that this confidence interval does not overlap the diagonal line and, thus, rejects equality of the forward and reverse rates. Ignoring confounding issues such as differential sampling of zygomorphic and actinomorphic flowers (whether by human bias or differential diversification), these results support the idea that there is a directional trend in the evolution of flower symmetry.

MODEL-BASED TESTS OF CORRELATED EVOLUTION

Having introduced model-based approaches for studying trait evolution along the branches of a specified tree, we can now return to the problem of determining if two traits have evolved independently. With these probabilistic models, it is possible to use a likelihood framework to test the hypothesis that two traits show correlated evolution. The most commonly used test is usually called the Pagel test, after Mark Pagel, who developed it in 1994. (Brook Milligan, who published the same test in the same year, rarely gets acknowledged, perhaps because he did not distribute software that others could easily use.) The Pagel test involves mathematically modeling correlated trait evolution.

As an example, let us consider the relationship between pollination mode (wind vs. animal; W vs. A) and showiness of flowers (inconspicuous vs. showy; I vs. S). We will explore the question of whether these traits are correlated, as might be expected given that showy flowers may be beneficial for attracting pollinating animals but may be disadvantageous under wind pollination (e.g., reducing wind flow or attracting flower predators).

The first step in the Pagel test is to use maximum likelihood to estimate the rate of evolution of the two traits under the null hypothesis that they evolve independently. Under this assumption, we can use a phylogeny to estimate four parameters, corresponding to the rates of gain and loss for each trait (Figure 10.10). This four-rate model is termed the independent model because the rate of evolution of one trait is not affected by the other trait.

Next, we examine the dependent model, in which the rate of change of one character depends on the state of the other. For example, we would allow the rate of evolving wind pollination to be different on lineages that have inconspicuous ($q_{AW|I}$) or showy ($q_{AW|S}$) flowers. As shown in Figure 10.11, we now have eight rates, each rate being conditioned on a state of the other character.

Because the independent model is less parameter-rich and nested within the dependent model, we can use a likelihood ratio test (Chapter 8) to evaluate whether the data justify the use of the more complex dependent model. If this test is not significant, then the simpler independent model is favored. If, however, the test is significant, then the dependent model is favored.

A few such studies have explored the relationship between animal pollination and showy flowers. For example, a study by Friedman and Barrett (2008) found strong support for the dependent model over the independent model. Animal-pollinated taxa tended to evolve showy flowers more often (and lose

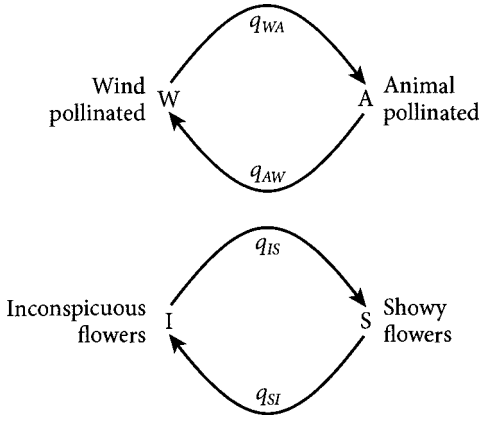


FIGURE 10.10 The independent model of evolution for pollination system and flower showiness. This model has four parameters, which describe the rate of evolution of each state of the two characters. Independence is shown by the fact that the rate of evolution at one character is unaffected by the state of the other character.

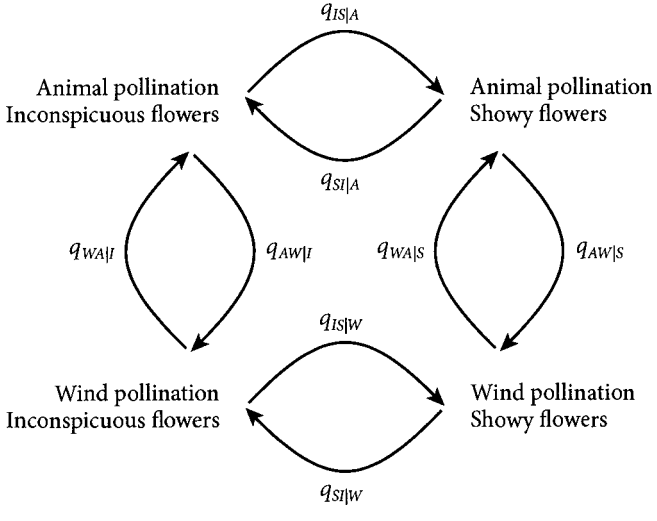


FIGURE 10.11 The dependent model of evolution for pollination system and flower showiness. This model has eight parameters, which describe the rate of evolution of each state of the two characters conditioned on each possible state of the other character. The "|" describes the condition; so, for example, $q_{IS|A}$ is the rate of going from inconspicuous to showy given that the lineage is animal-pollinated.

them more rarely) than do wind-pollinated taxa. This finding supports the hypothesis that showy flowers are often adaptations for attracting animal pollinators.

STOCHASTIC MAPPING

Even on a single tree there are an infinite number of possible histories of trait evolution that could give rise to the states at the tips. The maximum likelihood approach for studying trait evolution uses the substitution probabilities to integrate over all these histories. Similarly, Pagel's likelihood-based method does not commit to a single reconstruction of character history, but instead finds the model (and parameters) that best explain the data, while integrating over all nuisance parameters (e.g., exactly where the traits change state).

A Bayesian method, called ***stochastic mapping,*** approaches the problem differently. Rather than integrating over all character histories, stochastic mapping generates a sample of histories that should approximate the posterior distribution of histories. This sample can then be used to address questions about ancestral states, directionality, or correlated evolution. As with a posterior sample of tree topologies (Chapter 8), the frequencies of histories of different types should be proportional to their posterior probability. Therefore, by looking at the frequency of a specific outcome within the sampled histories, its posterior probability can be estimated.

Stochastic mapping generates the distribution of possible histories by sampling prior distributions for the necessary parameters: q_{10}, q_{01}, and a tree length statistic. The latter indicates the total number of changes in character state expected over the whole tree. You may notice that these are the same rate parameters used in likelihood methods. Indeed, stochastic mapping is founded on the same underlying model of trait evolution and uses that model to simulate character histories or "realizations" up a tree.

The simulation proceeds in three steps. First, the relative posterior probability of state 0 and state 1 at each internal node is calculated. Next, one assignment of states to internal nodes is picked from this distribution, meaning that for each branch we know its starting and ending state. These two steps are included to decrease the practical and computational challenge of simulating a tree that terminates in the observed tip states by breaking the problem into smaller pieces (each individual branch). Finally, based on the selected rates,

evolution is allowed to occur up each branch individually, repeating the procedure until each branch ends at the required ending state. At this point we would have a single realization of the character history.

In each realization, stochastic mapping provides a complete picture of all the changes on the tree that occurred during the simulation (Figure 10.12). This is somewhat like parsimony, where changes are associated with particular branches of the tree, except that stochastic mapping specifies exactly where along each branch changes occurred. Also, in contrast to parsimony, stochastic mapping allows for more than one change along a single branch.

Figure 10.12 shows two realizations of the history of swords in the *Xiphophorus* fishes from Figure 10.7. In order to estimate the posterior distribution of character evolution histories, we would repeat the realizations many times, and in so doing, obtain a posterior distribution for the number of changes in each

FIGURE 10.12 Stochastic mapping of presence/absence of swords in *Xiphophorus* fish. Two realizations of character history are shown. A full analysis would generate thousands of such realizations to determine the frequencies of different outcomes.

direction and the states at each node. Additionally, we could repeat these simulations across multiple trees from a Bayesian MCMC search (Chapter 8) in order to incorporate uncertainty in the phylogeny.

In addition to estimating ancestral states for a single character, stochastic mapping can be used to test for correlated evolution between two characters (call them 1 and 2). To do so, we generate a large number of simulations to approximate the posterior distribution of histories for each trait. Then we overlay the history of both characters and look at the total amount of the history that has each of the possible combinations of the two characters (state 0 for character 1 and state 0 for character 2; state 0 for character 1 and state 1 for character 2, etc.). For example, if transitions from 0 to 1 in character 1 are correlated with changes from 0 to 1 in character 2, we would predict that the 0 states and the 1 states for each character will co-occur more often than expected by chance. While not commonly done, it would also be possible to do a model-based version of the concentrated changes test. For example, we could look at the proportion of simulations that have more gains of character 1 on lineages that are in state 0 than in state 1 of character 2 and thereby assess whether character 2 affects the evolutionary dynamics of character 1. The statistical methods used for significance testing with stochastic mapping are described in more detail in the recommended Further Reading.

THE EVOLUTION OF CONTINUOUS TRAITS

The preceding discussion of trait evolution has dealt with discrete traits: characteristics that are defined to have two or a few possible states. A head crest can be treated as being present or absent, flowers may be scored as either pink or white, and the nucleotide at position 63 in the ß-hemoglobin gene can be an A, C, G, or T. In many evolutionary contexts it is reasonable to think in terms of discrete traits.

For some traits, however, the observed variation among organisms cannot be broken up into a series of discrete states whose boundaries are sharp. Consider body size, for example, which may be influenced by many genes. Even if each gene that potentially affects body size has a discrete effect, the summation over the many body-size-affecting genes results in the trait evolving more smoothly. Instead of transitioning in quantum jumps, as assumed for discrete characters, body size will tend to gradually increase or decrease during

evolution. Thus, for traits like body size it is more accurate to model their evolution as a ***continuous*** trait. Instead of being divided into discrete character states, each organism has a particular value of that trait (e.g., 1200 g) and any population or species has a mean size and variation around that mean (e.g., 1200 g ± 12 g).

As with discrete traits, the discovery of an evolutionary correlation between two continuous traits can provide insights into adaptive evolution. For example, discovering a correlation between animal body size and life span has important implications for theories of the evolution of aging.

You might wonder whether we can detect an evolutionary correlation between traits without worrying about phylogeny. After all, if two continuous traits were correlated during evolution, would we not see that correlation simply by looking at the values of the two traits found in the living species? This logic is flawed. In the same way that we cannot detect correlated evolution using a contingency table of discrete traits (Figure 10.3), we also cannot infer correlated evolution from correlations in the values seen in the tips. This is illustrated in Figure 10.13, which shows the values of two hypothetical continuous traits for 20 species. Looking only at the tips, the pattern is highly suggestive of a positive correlation between traits X and Y. However, when interpreting this scatter plot we need to remember that species are not independent data points, because they share significant parts of their history in common. This can

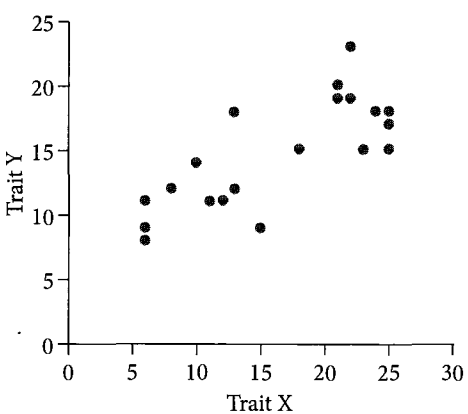


FIGURE 10.13 Scatter plot showing an apparent positive correlation between two continuous traits. Each dot represents a species with a particular value of traits X and Y.

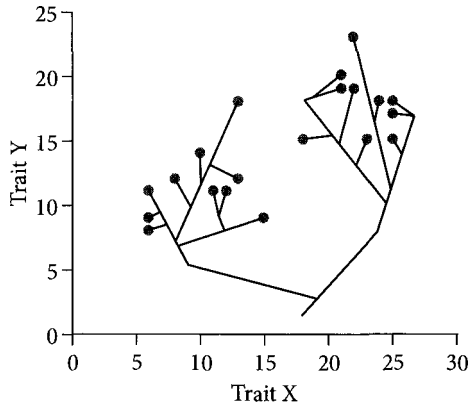


FIGURE 10.14 The trait values shown in Figure 10.9 with the evolutionary relationships shown. The apparent tip correlation between X and Y is driven mainly by two clades, one with high X and high Y, one with low X and low Y. Given this phylogeny, the apparent evolutionary correlation between X and Y ceases to be convincing.

be seen by overlaying the phylogenetic tree for these 20 species on the graph. As Figure 10.14 shows, what appeared to be a broad-scale correlation boils down to two clades, one with generally lower values of X and Y, and one with generally higher values. Within these two clades there is no pattern of correlated evolution. Thus, the inference of a correlation between X and Y was really just based on two data points—hardly a solid foundation for drawing conclusions about correlated evolution!

We cannot use the same methods to study continuous traits as we use to study discrete traits. In the case of discrete traits, we generally assume that a lineage jumps abruptly between states. Maximum likelihood approaches allow for variation in the rate at which these jumps happen, while parsimony approaches assume that jumps happen relatively rarely. Our assumptions about how continuous traits change are inherently different. We expect trait values to be evolving to some degree at almost every point in time. Instead of changing by rare jumps, we expect descendants to have a trait value similar to but not identical to those of its recent ancestors.

The best way to visualize the evolution of a continuous trait along an evolving lineage is to start by assuming that traits evolve neutrally in a Markovian manner. This means that mutations that move a lineage back toward a previous

trait value are just as likely as ones that move the lineage toward a new value. Consider a living species whose mean body size is 1200 g, but which evolved from an ancestral lineage that one million years ago had a body size of 1000 g. Unless we happen to know that there has been consistent evolutionary pressure favoring larger organisms, we would have to say that one million years from now the trait value is equally likely to have decreased back toward 1000 g as to have increased toward 1400 g. Thus, the evolution of a continuous trait is best visualized as a random walk through a space of trait values.

To help you visualize the evolution of continuous traits, we can use a metaphor. Imagine an American football field in which the yardage lines represent time and the left-to-right dimensions (distance from the midline) represent the value of a trait: body size or maximum running speed or leaf length, for example. The closer you are to the left side of the field, the lower the value of the trait, and the closer you are to the right side of the field, the higher the value of the trait.

Imagine a player starting at one goal line and then walking blindfolded toward the other end of the field. Because the player might inadvertently drift left or right, it is likely that his side-to-side position (i.e., the trait value) when he reaches the other goal line will not be the same as when he started. For example, if he started at the midline (the center of the goal line), he might end up 15 yards left (−15) or five yards right of the midline (+5). However, since the field is level, the player will be as likely to end up to the right or left of his starting position.

The process may be modeled as Brownian motion. This refers to the motion of small particles being bumped by random hits from water or air molecules around them, first reported by Robert Brown in the early nineteenth century. Each time a player walks down the field the actual path is random. So if multiple players all started at the same position, they would probably cross the other goal line at different positions. Figure 10.15 shows an example in which five players are started at the midline. The expectation is that the mean position of the players once they get to the goal line should be equal to the starting position. Furthermore, given Brownian motion, the positions of individual players will be normally distributed about this mean, with a variance that is linearly proportional to the distance that the players have walked (analogous to the time that traits have evolved). While this is hard to see with just five runners, if there were a sufficient number of runners, you would see that they would gradually spread out from one another as they proceeded downfield.

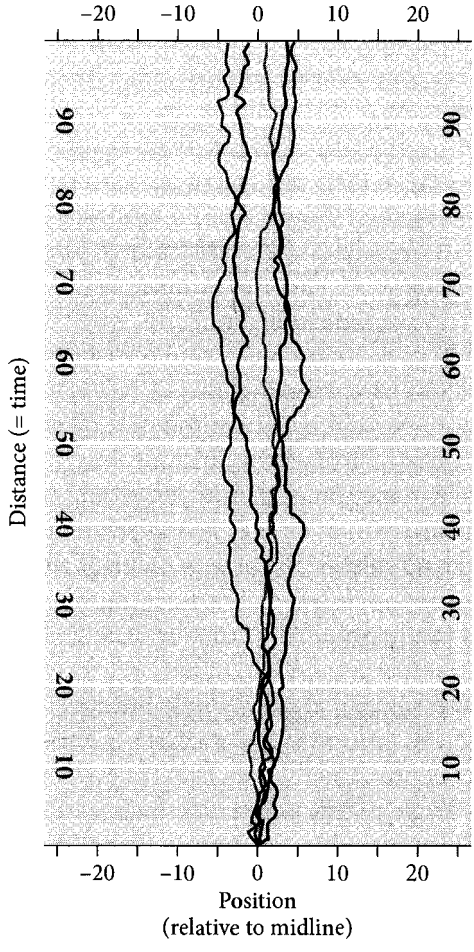


FIGURE 10.15 Paths of five blindfolded football players as they cross a football field, as an illustration of Brownian motion trait evolution. Five players independently start from the center of the goal line (midline) at one end of the field and run downfield. While the mean position at which the five players cross the other goal line will be close to the midline, some players will, by chance, have deviated to the right or left.

Under a Brownian motion model of trait evolution, the rate at which a lineage changes to higher or lower values is an intrinsic property of a trait. One can imagine that this rate of evolution is determined by the number of variable genes that affect the trait, the trait's sensitivity to fluctuating selective pressures, and so on. Consequently, traits can differ in their rate of evolution. By analogy, football players might vary in their ability to keep a straight line. If Figure 10.15 represents typical players, then Figure 10.16 might represent players who are less able to keep to a straight line when blindfolded. The fundamental difference between these two figures is that the variance in player position increases more rapidly as a function of time in Figure 10.16. Applying this to continuous trait evolution, the concept of evolutionary rate refers to the rate at which variance is expected to grow as a function of time.

Now let us think about our ability to infer the starting position of a runner by observing where he is when he crosses the finishing line. If we observed multiple runners that had started at the same position, we would be confident that they had all started close to the mean of their finishing positions. The same would be true if we observed a single runner: the best estimate of where he started would be the same position that he finished. The uncertainty around this estimate would be determined by the product of the distance he traveled and the rate at which he tended to shift right and left. Likewise, according to the Brownian motion model of trait evolution along a single lineage, the best estimate of a trait's value in the past is its current value, and the variance around this estimate is given by the product of time and the trait's rate of evolution.

The metaphor of a single player walking across a football field is analogous to evolution in a single, unbranching lineage. In order to extend this metaphor to branching phylogenies, we need to imagine that one player starts walking across the field, but then stops at some point (20 yards in Figure 10.17a) and tags two teammates, who start where they were tagged and conduct their own walk downfield. The two tagged players start at the same distance from the midline, analogous to the way that two descendant lineages would be expected to initially have the trait value of their common ancestor. However, as time progresses the two tagged players are likely to diverge.

As shown in Figure 10.17a, by the time the two players reach the far goal line, they can have very different trait values. This is especially true when two lineages have evolved independently for a long time. If, instead, the lineages have diverged much more recently, as in Figure 10.17b, the difference in their

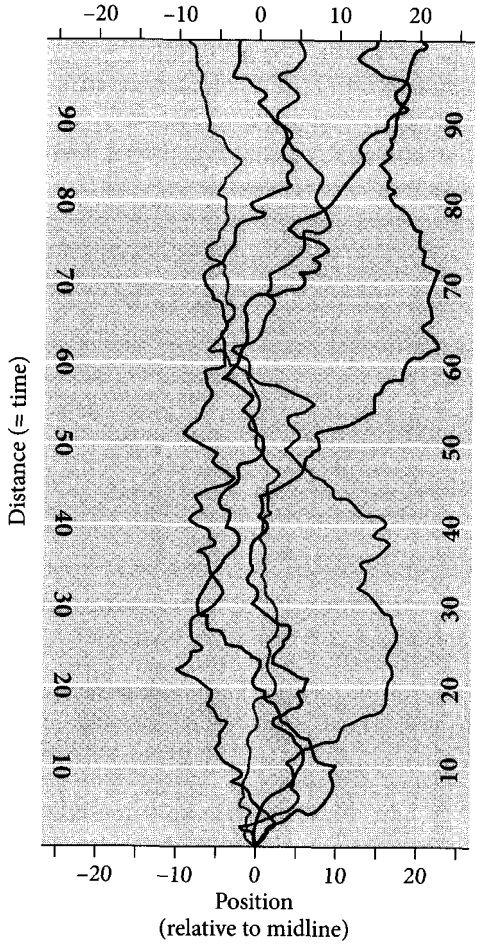


FIGURE 10.16 Paths of five different football players whose rate of deviating left or right when blindfolded is higher than those in Figure 10.15. Again, the average position of players at the end of the run is close to the midline. However, relative to the players in Figure 10.15, there is a greater variance in the players' position at the other goal line.

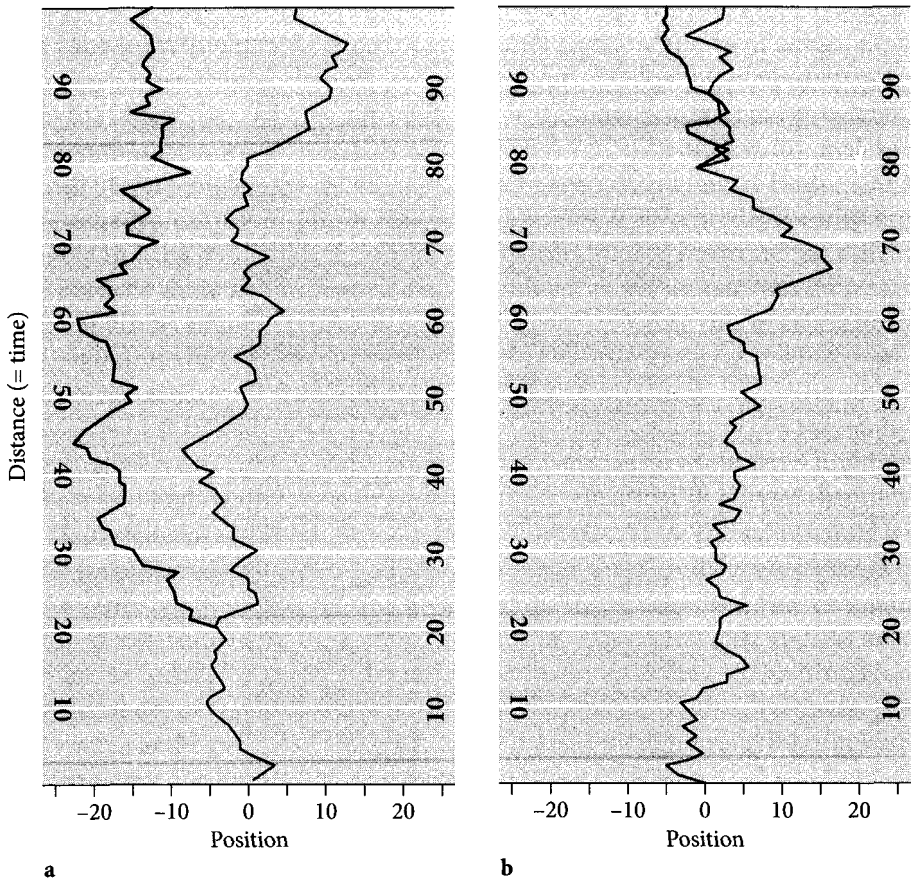


FIGURE 10.17 Paths of football players in a branched relay, a metaphor for continuous trait evolution along the branches of a phylogenetic tree. When a lineage branches, each descendant begins with the same trait value but, over time, the trait values of the two descendants can deviate. In the first case (a), the two players running the second leg have had longer to diverge in position than in the second case (b), where the handoff happened close to the end of the race.

final trait values is expected to be less. Given this way of thinking, the extent to which two taxa have similar trait values is strongly influenced by how recently they shared common ancestry. This is an important general principle of trait evolution. While distant relatives will sometimes converge on similar trait values, very close relatives must have similar trait values because they recently

diverged from a common ancestor. Thus, we have an asymmetry in inference that goes some way toward explaining why similarity and recency of common ancestry are distinct and potentially conflicting criteria (Chapter 5): similar species may or may not share recent common ancestors, but species that do share a recent common ancestor will tend to be similar.

If we assume evolution by Brownian motion, it is possible to use a tree with branch lengths to obtain an estimate of trait value for an ancestral node as well as a measure of uncertainty around that estimate. Such ancestral state reconstruction can be accomplished using a method called ***squared-change parsimony.*** To

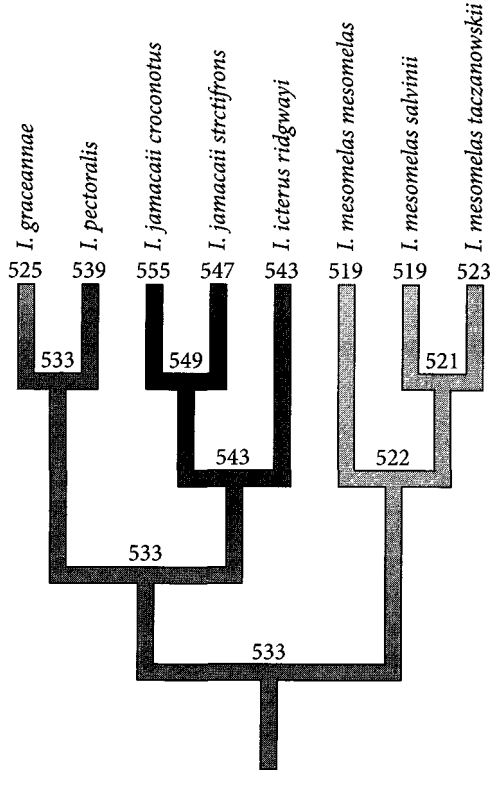


FIGURE 10.18 Squared-change parsimony reconstruction of breast color in orioles. The breast color (in wavelength units) of the eight living taxa is shown at the tips, and the best estimate of the wavelength at each internal node is also shown above the node. Branches are shaded based on their inferred wavelength. Based on Hofmann et al. (2006).

give one example, Figure 10.18 shows an analysis that used a phylogenetic tree of orioles to study the evolution of breast color (expressed as wavelength). As you can see, the point estimate of a node's trait value tends to lie between the values of the three surrounding nodes (two descendant nodes and one ancestral node). However, the confidence intervals around a node's trait value (not shown) can be quite broad, and can exceed the values seen at surrounding nodes.

The Brownian motion model of trait evolution implies that the variance in trait value grows indefinitely, meaning that the trait range should ultimately vary from minus infinity to plus infinity. Clearly, this is an unrealistic assumption, which is one reason why other models of trait evolution have been developed. The most popular is the ***Ornstein-Uhlenbeck (OU) model,*** which allows for some element of stabilizing selection that keeps traits close to an optimal value. OU assumes that, the further lineages deviate from their optimum, the more forcibly they are pulled back to it. Returning to the football field metaphor, imagine that the football players are tethered to a rubber band that is attached to a rail down the midline of the field. This would not stop them wandering, but would tend to ensure that they don't wander off the field entirely.

The OU process is mathematically more complex than Brownian motion, adding another parameter, the strength of stabilizing selection. If the strength of stabilizing selection is set to zero, the model behaves exactly like Brownian motion. This shows that the Brownian motion model is a special case of the OU model. While the OU model is more complex, it is often favored because it is more biologically realistic.

DETECTING CORRELATED EVOLUTION OF CONTINUOUS TRAITS

Up until now, we have only considered the evolution of one continuous trait at a time. When two traits are evolving along a lineage, it is possible that they affect one another's evolution. Such correlated evolution can be visualized by imagining that the pair of players running the equivalent leg of the relay are linked by a rubber band. This band means that when one player veers to the left he tends to pull the other player to the left as well. In biological terms, we expect to see that species with high values for trait 1 will also have high values for trait 2 if the two traits are positively correlated (if they are negatively correlated, we just need to imagine reversing the scale for one trait). The length of the band

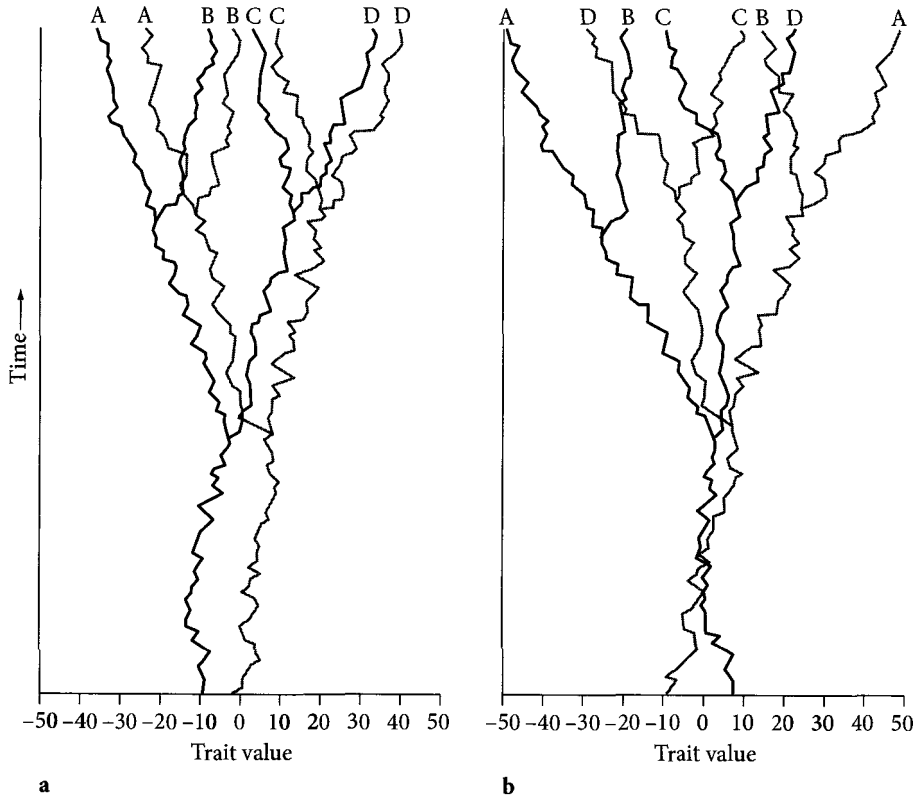


FIGURE 10.19 Evolution of two continuous traits along the same four-taxon tree by Brownian motion. In panel (a) the two traits are evolutionarily correlated, meaning that when one trait increase the other trait tends to increase also. For contrast, panel (b) shows traits that are evolving independently.

can be seen as a measure of the strength of trait correlation. Figure 10.19a illustrates the evolution of two correlated traits along the same four-taxon tree. In Figure 10.19b the topology and branch lengths are the same, but the two traits are not correlated.

There are multiple phylogenetic comparative methods that may be used to assess whether two continuous traits have undergone correlated evolution. The simplest to explain is ***phylogenetic independent contrasts.*** The first step is to estimate the values of each trait at each ancestral node based on the values of

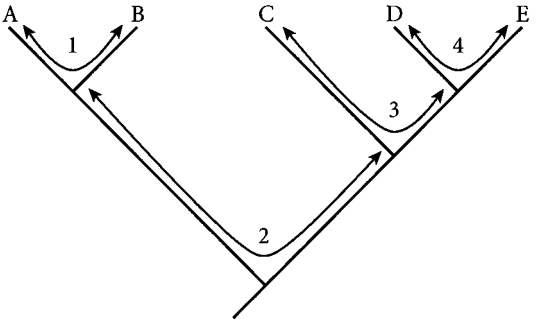


FIGURE 10.20 The four independent contrasts on a five-taxon tree. A comparison of trait values between sister groups is called an independent contrast. These are useful for comparative analyses because breaking a tree into independent contrasts ensures that information on each branch only impacts the analysis once.

that node's descendants. This is usually done by assuming Brownian motion. Then we can break the tree up into a set of independent sister-group comparisons, called ***contrasts.*** Figure 10.20 shows a rooted tree with five taxa and four independent contrasts (there is always one fewer contrast than taxa). You will see that the contrasts do not overlap—no part of the tree is used in more than one contrast. This is what is meant by *independent* contrasts. By dividing the tree into nonoverlapping pieces, any correlations that are due to phylogenetic structure alone are eliminated. As a result, we can calculate the contrasts for the two traits and see if they are correlated using standard statistical methods. You can consult the recommended Further Reading to learn how this is done while taking account of uncertainty in the ancestral states.

To provide an example of the application of independent contrasts, consider the question of whether body size and home range are correlated in mammals. We have reason to imagine these two traits might be correlated, since large animals need larger range sizes in order to obtain enough food to support their body mass. Using both ungulates (hoofed mammals) and carnivorans, Garland et al. (1992) isolated the independent contrasts for these two traits and plotted them out. As shown in Figure 10.21, there is a significant positive correlation between body mass and range size in both groups. This is compatible with body size evolution being shaped by available range size or, conversely, with body size evolution resulting in changes in range size.

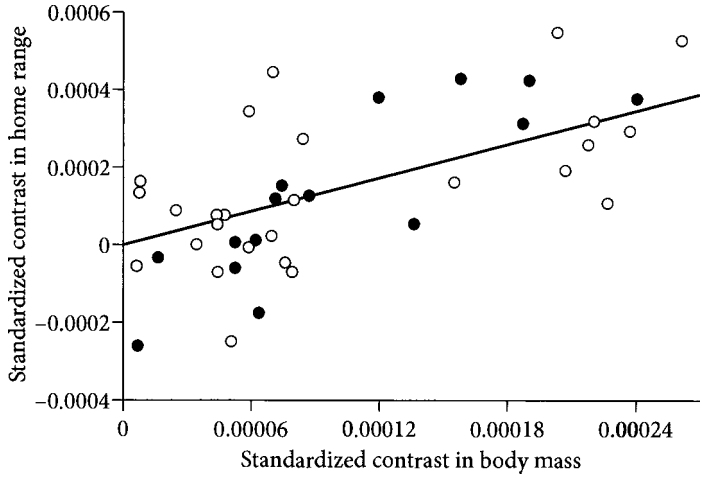


FIGURE 10.21 **A plot of standardized contrasts of body mass and home range in ungulates (open circles) and carnivorans (solid circles).** These traits show a significant phylogenetic correlation in both mammal groups. Adapted from Garland et al. (1992).

The development of phylogenetic independent contrasts in 1985 had a profound effect on the study of correlated evolution in continuous traits. Since that time it has ceased to be acceptable to make cross-species comparisons in such fields as ecology and functional morphology without making an effort to accommodate phylogenetic structure. However, because the phylogenetic independent contrasts method assumes that traits evolve via Brownian motion, an assumption that is frequently violated, several alternative comparative methods have been developed.

The most widely used alternative to independent contrasts is ***phylogenetic generalized least squares*** (PGLS). This method can implement the OU model, can accommodate measurement uncertainty in the tip data, and can be scaled up to conduct multiple regression analysis. The result of PGLS analysis is an estimate of the slope of the regression of one trait on another and an indication of whether this slope is significantly different than zero.

As an example of the application of PGLS, Clusella-Trullas et al. (2008) used the method to test for an association between skin reflectance of lizard species and the mean annual radiation they experience. They found a significant

positive association between these two traits, meaning that there has been an evolutionary correlation between skin reflectance and sun exposure in lizards: lineages living in more sunny locales reflect more light and those living in less sunny habitats absorb more light. Such a conclusion shows that, by using a phylogenetic approach, we can find strong evidence to support the role of selection in shaping the evolution of interesting traits such as skin pigmentation.

FURTHER READING

Phylogenetic approaches to the study of adaptation: Gould and Vrba 1982; Baum and Larson 1991; Harvey and Pagel 1991

Concentrated changes test: Maddison 1990

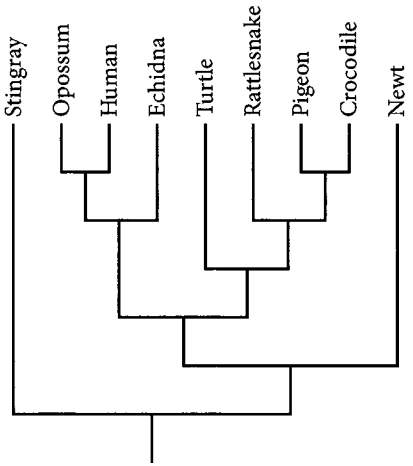
Maximum likelihood methods for ancestral state reconstruction of discrete characters and correlated evolution: Pagel 1994; Schluter et al. 1997

Stochastic mapping: Huelsenbeck et al. 2003

Modeling the evolution of continuous traits: Felsenstein 1985b; Garland et al. 1992; Martins and Hansen 1997; Butler and King 2004

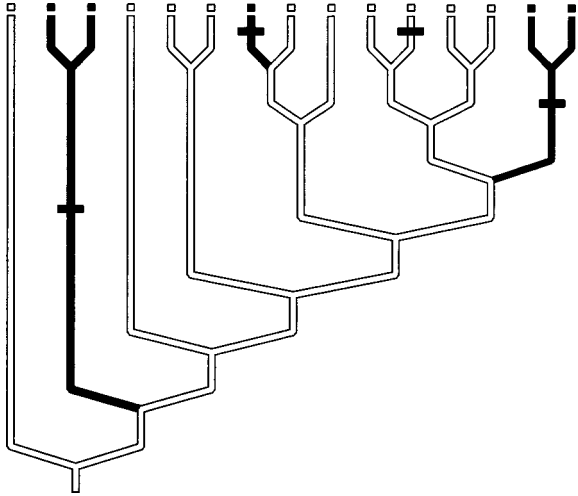
CHAPTER 10 QUIZ

1. It is known that having separate bones in the skull is beneficial in humans by allowing the skull to deform during birth. The tree shows selected vertebrates. Of these, only humans give live birth of a large-headed infant; the others lay eggs. Also, all taxa except stingrays have a skull made of separate bony plates. Assuming parsimony, what can we conclude about the status of separate skull bones?
   a. Separate skull bones in vertebrates are adaptations to egg laying.
   b. Separate skull bones in humans are adaptations to live birth.

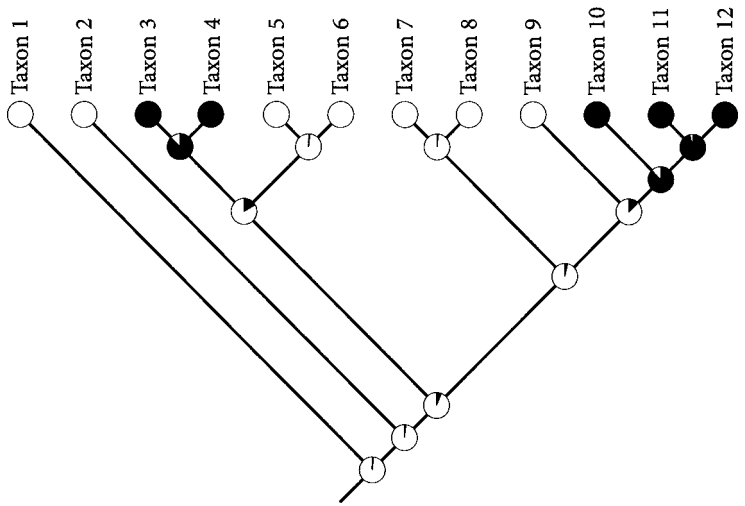


c. Separate skull bones in humans are exaptations to live birth.
d. Live birth in mammals is an exaptation to separate skull bones.
e. None of the above can validly be inferred.

2. Consider this phylogeny, which traces the history of sexual dimorphism in a hypothetical group of birds, with black lineages being dimorphic (different male and female forms) and white lineages being monomorphic. Colorful plumage has arisen four times in the group, as marked by the black bars. What does this pattern suggest about plumage coloration and dimorphism?
   a. Colorful plumage is just as likely to arise in monomorphic lineages as in dimorphic lineages.
   b. The evolution of colorful plumage is concentrated in dimorphic lineages.
   c. Changes in coloration are randomly distributed across the tree.
   d. The combination of colorful plumage and dimorphism leads to speciation.
   e. Plumage and sexual dimorphism are probably evolving independently.

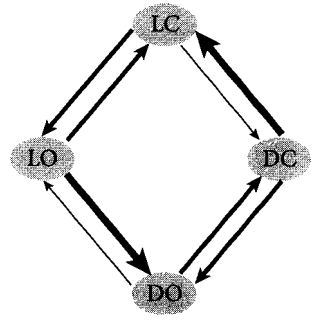


3. The tree below shows a maximum likelihood reconstruction for thorns in a group of 12 taxa (white = no thorns, black = thorns). The pie graphs at each node display the relative likelihood of each state (no thorns or thorns). What can we say given this reconstruction?
   a. The common ancestor of the group most likely had thorns.
   b. Thorns have evolved exactly twice in the group.
   c. There is a small probability that taxon 10 lacks thorns.
   d. Thorns have never been lost in the history of the group.
   e. The thorns in taxa 3 and 10 have a low, but nonzero, probability of being homologous.



4. There are many clades of flowering plants that have both fleshy and dry fruits. Imagine we surveyed the fruit type in one such group and used the phylogeny for this group to estimate the rates (q_{fd}, q_{df}) for transitions between the fleshy (f) and dry (d) types. If we estimated that q_{fd} was 16.1 ± 0.1 and q_{df} was 0 ± 0, what would that imply about the evolution of fruit type?
   a. The two rates, q_{fd} and q_{df}, are statistically different from each other.
   b. There is an evolutionary trend away from fleshy fruits.
   c. Fleshy fruits and dry fruits are not expected to be equally frequent among the tips in our phylogeny.
   d. The evolution of dry fruits is effectively irreversible.
   e. All of the above.

5. Many animals living in open habitats like the North Pole have lighter fur than those in closed habitats like forests. Imagine we used maximum likelihood methods to test if fur color and habitat type showed correlated evolution. The figure shows the four possible combinations of fur (L = light, D = dark) and habitat type (O = open, C = closed) and the estimated rates of transition between each of the four states. Assuming that the thickness of an arrow indicates the relative rate of a particular transition, what might we conclude?



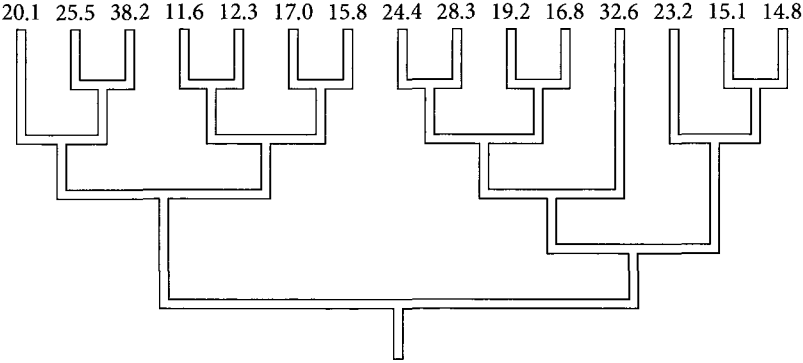
   a. Gains of dark fur are much more likely than losses.
   b. There is significant directionality in the evolution of habitat type.
   c. Transitions from light to dark fur are more likely in open than in closed habitats.
   d. Transitions from dark to light fur are less likely in closed than in open habitats.
   e. All of the above.

6. Following the example in Question 5, what might we infer about the evolution of the two traits?
   a. Changes in fur color depend on habitat.
   b. Habitat transitions depend on fur color.
   c. The two traits evolved independently along the tree.
   d. The ancestral state was light fur and an open habitat.
   e. The ancestral state was light fur and a closed habitat.

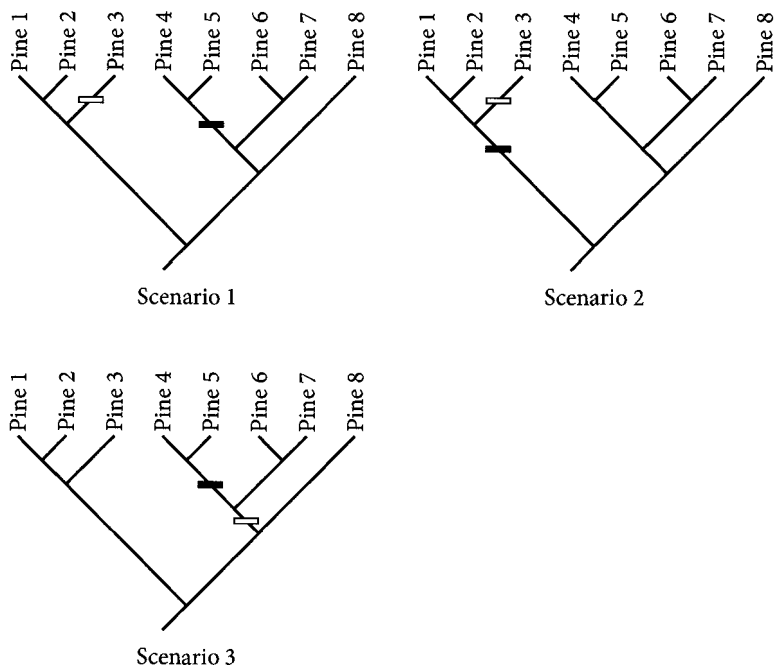
7. The difference between the Brownian motion (BM) and Ornstein-Uhlenbeck (OU) models of evolution is that . . .
   a. . . . the OU model is simpler than the BM model.
   b. . . . only OU models have an expected value for the trait after a period of time.
   c. . . . rates of trait evolution can vary only under the BM model.
   d. . . . the expected variance in trait values increases linearly with time in BM but not OU.
   e. . . . the BM model is usually more biologically realistic than the OU model.

8. In the radiation of Darwin's finches on the Galapagos Islands, the differences in beak size are thought to reflect adaptation for different-sized seeds. Imagine we measured beak size and seed diameter for 15 species of finches and found the pattern shown (the upper number for each tip is beak size and the lower number is seed size). We calculate independent contrasts for such a data set and estimate the correlation between the contrasts to be 0.95 ($P < 0.001$). How would we interpret this result?
   a. There is an evolutionary trend toward increasing beak size and seed size.
   b. Increases in seed size lead to decreases in beak size.
   c. Larger beak sizes are correlated with higher values for seed diameter.
   d. There is significant support for the hypothesis that seed diameter and beak size evolved independently.
   e. There could be an evolutionary correlation between these traits, but, to be sure, we need to take account of the structure of the phylogeny.

Beak size	5.2	6.7	10.1	1.1	1.4	3.2	2.0	5.5	6.3	4.8	4.6	7.0	5.0	2.2	2.5
Seed size	20.1	25.5	38.2	11.6	12.3	17.0	15.8	24.4	28.3	19.2	16.8	32.6	23.2	15.1	14.8

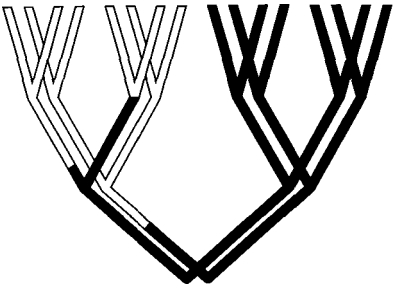


9. Some biologists have hypothesized that spines on pinecones are adaptive for deterring predation by birds. Whereas squirrels pry off the scales to get to the seeds inside, birds reach down into the scale to get to the seeds and could be blocked by a large spine. The figure shows a phylogeny for eight pine species and three possible scenarios for the origin of scale spines (black bars) and bird predation (white bars). Which of the scenarios would support this adaptive hypothesis?
   a. 1 b. 2 c. 3 d. Both 2 and 3 e. Both 1 and 2



10. Reconsider the hypothesis that pinecone spines are adaptations to reduce seed predation by birds. Suppose you confirmed that the presence of spines reduces seed predation, for example, by showing experimentally that removing of spines increases seed predation rates. Also assume that pinecone spines evolved before the evolution of seed-eating birds (scenario 2 in Question 9). Provide a reasonable evolutionary narrative to reconcile these findings.

11. Stochastic mapping of two independent traits can be used to test for correlated evolution. When two character states (e.g., 0 for character 1 and 0 for character 2) co-occur along a branch more often than expected, this is evidence for correlated evolution. The figure, adapted from Huelsenbeck et al. (2003), shows two sets of trees. Each set shows a realization of stochastic mapping for two characters (L, U). The tables show the amount of overlap (expected and observed for each pair of states). For example, in the top set of trees, the observed overlap of the 0 state for character L and the 0 state for character U (colored as white on the tree) accounts for 35% of the branch length along the tree. Which set of trees (top or bottom) shows greater support for correlated evolution of L and U? How do you know?

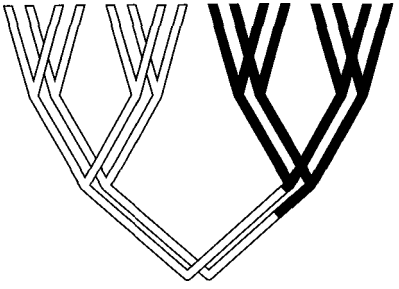


Observed

		U	
		0	1
L	0	0.35	0.11
	1	0.00	0.54

Expected

		U	
		0	1
L	0	0.16	0.30
	1	0.19	0.35



Observed

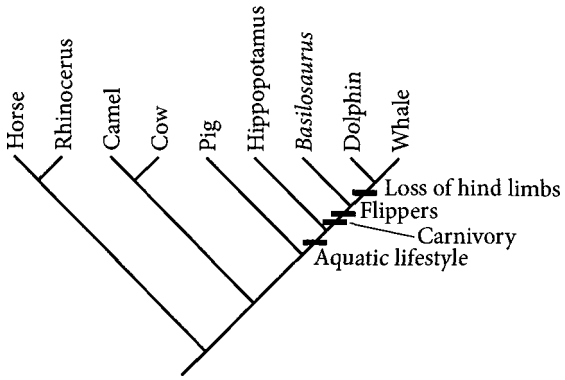
		U	
		0	1
L	0	0.55	0.00
	1	0.02	0.43

Expected

		U	
		0	1
L	0	0.31	0.24
	1	0.26	0.19

12. Suppose you could travel through time and examine leaf length through the history of an evolving plant lineage. Two million years ago the leaves averaged 10 cm in length. One million years ago the leaves averaged 8 cm in length. In the present the leaves are on average 6 cm long. If leaf length were evolving according to a Brownian motion model, what can you say about the expected leaf length one million years into the future? Is there a reason to doubt that this trait is evolving according to the Brownian motion model of trait evolution?

13. The tree shows the relationships between cetaceans (dolphins and whales), a fossil relative (*Basilosaurus*), and other ungulates (pigs, cows, etc.). Assuming that this is the correct tree, provide a brief evolutionary chronicle and a brief evolutionary narrative for this group.



14. Why is correlated trait evolution considered evidence for adaptation?

15. Explain why it is important to take phylogeny into account when testing for relationships (e.g., correlations) between continuous traits.

16. There are multiple methods for testing for correlated evolution between discrete traits, including the concentrated changes test, the Pagel test, and stochastic mapping. Are there any scenarios in which one would favor the parsimony-based concentrated changes test over the two model-based approaches?

CHAPTER ELEVEN

Using Trees to Study Space, Time, and Evolutionary Diversification

As the previous chapter showed, phylogenetic trees provide a framework for studying the evolution of traits. Not only do they help us to make statistical inferences about the branches on which certain traits arose, they also give us tools for exploring the action of selection in explaining the distribution of traits among living species. In this chapter, we examine how the methods introduced in relation to trait evolution may be applied to some more challenging problems: explaining the geographic distribution of living (and extinct) species, exploring how species affect one another's evolution, understanding how ecological communities are assembled, estimating the absolute time frame of evolution, and studying why some clades contain more lineages than others. While this is a somewhat heterogeneous group of topics, they all require us to clearly visualize the evolution of lineages in space and time while being buffeted by interactions with distantly and closely related organisms.

RECONSTRUCTING DISPERSAL HISTORY

As discussed in Chapter 4, geographical location is a heritable trait. In the same way that lineages drift through trait space (the metaphorical football field from Chapter 10), they also move through geographical space. However, geographical space is more complicated because the regions that an organism can potentially live in are scattered over the globe and constantly in flux. A population may be surrounded by inhospitable environments, but a perfectly habitable site may be available some distance away. Depending on the dispersal ability of the organisms, it may require a very rare event for new, distant sites to be successfully colonized.

As discussed in Chapter 2, the rarity of long-distance dispersal explains why taxa tend to live close to their immediate relatives, a fact that led both Darwin

and Wallace to the theory of descent from common ancestry. Interestingly, it is the very rarity of long-distance dispersal that makes the phenomenon easy to study using phylogenetic methods. So long as the rate of dispersal from one isolated habitat to another is lower than the rate at which lineages diversify within a habitat, we are able to use parsimony to obtain a reasonable reconstruction of the migration of lineages around the globe.

Figure 11.1 illustrates this with an analysis focused on Hawaiian plants in the family Lobeliaceae. These plants are morphologically diverse, so diverse

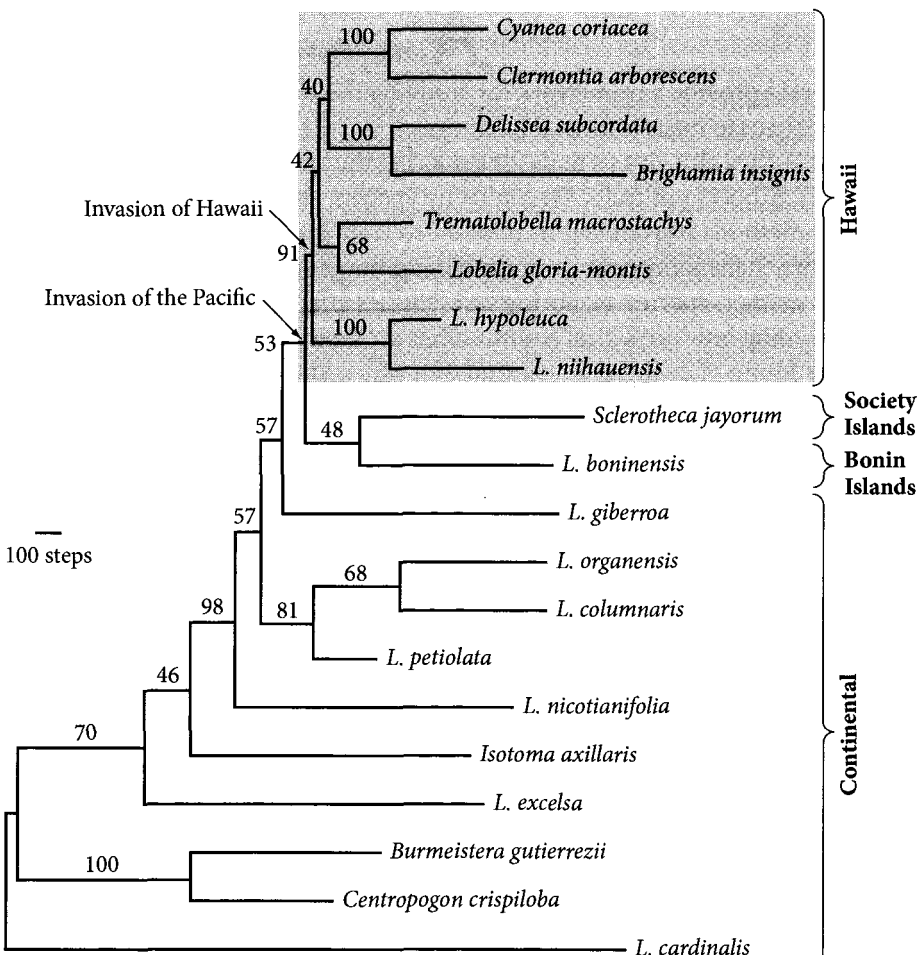


FIGURE 11.1 Phylogenetic reconstruction of plants of the Lobeliaceae. This result suggests (by parsimony) a single invasion of Hawaii. Adapted from Givnish et al. (2009).

that they were traditionally placed in six genera that were thought to represent multiple, separate dispersals to Hawaii. However, molecular phylogenetic analyses found that the Hawaiian lobeliads constitute a single clade that is embedded within the worldwide genus *Lobelia*. By parsimony, we can explain this tree with a single invasion of Hawaii. Furthermore, the closest relatives of the Hawaiian taxa seem to live on other Pacific islands. Thus, these data also suggest a single invasion of the Pacific. However, based solely on parsimony, it is unclear whether the invasion of the Pacific began in Hawaii, with subsequent dispersal to other Pacific islands, or vice versa.

If the rate of migration is high, as might happen with highly mobile species or when similar habitats occur in close proximity, migrations in space can occur so frequently that it becomes difficult to reconstruct movements with confidence. An example is provided by the cat family (Figure 11.2). A phylogenetic analysis suggests some geographically constrained clades, for example, the ocelot clade, which is restricted to South America. However, the broader

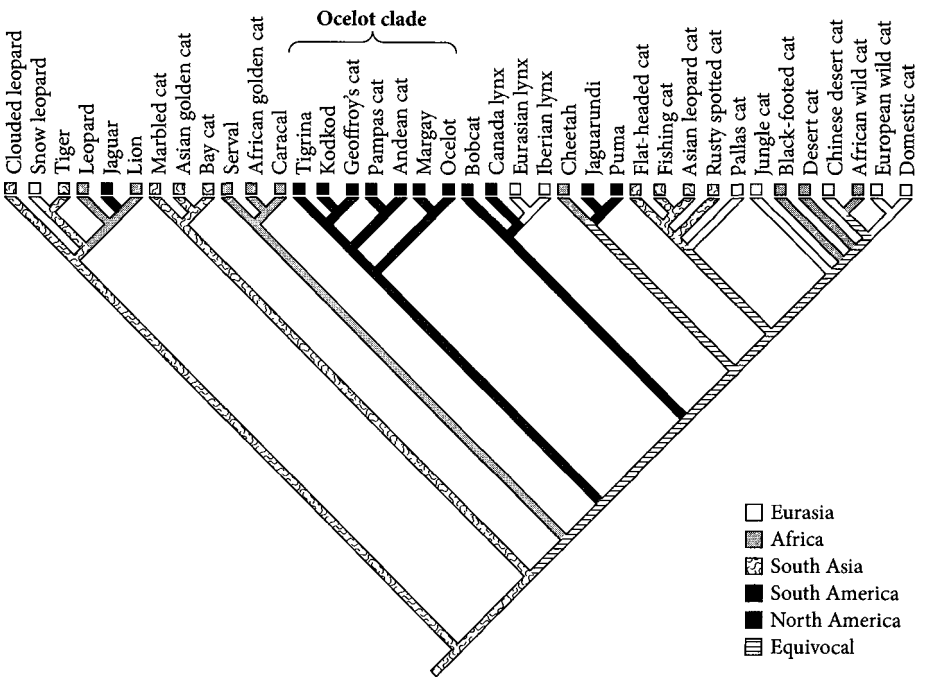


Figure 11.2 **Parsimony reconstruction of the biogeographic history of cats.** Because there have been many dispersal events, it is difficult to convincingly reconstruct cat migrations using parsimony. Adapted from Johnson et al. (2006).

pattern of cat evolution appears to have involved many intercontinental migration events. In such cases, the order of migrations becomes a source of uncertainty. This can be seen by applying equally weighted parsimony, which finds many equally parsimonious reconstructions. Furthermore, because of the high rate of dispersal, we should not have high confidence in the biogeographic conclusions arrived at by the application of parsimony. For example, while parsimony suggests that the last common ancestor of lions and leopards lived in Africa, the great migration ability of large cats means that two independent invasions of Africa cannot be ruled out.

VICARIANCE

The challenge of reconstructing biogeographic history is made even more difficult by the fact that climate change and continental movements are constantly altering the spatial distribution of suitable habitats. New suitable habitats may open up, and areas that were once habitable may cease to be so. The latter phenomenon can result in the fragmentation of geographic ranges, which can play an important role in shaping the current distributions of taxa.

Imagine a widespread species, composed of many local populations that are glued together by occasional gene flow. If the environment changes, this once contiguous population may be subdivided into two or more separate population lineages. These newly separated lineages can then evolve independently, ultimately forming distinct species. The resulting species are likely to fill similar ecological roles in their respective communities. For instance, if a shade-loving grass species was split into two by continental drift, it would tend, at least initially, to give rise to two similarly shade-loving species. Thus, the phenomenon of a widespread population being fragmented and giving rise to multiple descendants is termed ***vicariance***, from the Latin "vicarius" meaning substitute or successor.

In cases where species ranges have been shaped more by vicariance than by dispersal, the topology of the tree documents the order in which geographic space was fragmented. Figure 11.3 illustrates a hypothetical example in which a broadly distributed ancestral species is divided into a western and an eastern half by the cutting of a river canyon. Later, both the western and eastern populations are split into northern and southern halves, for example, by drying of a central strip of terrain. Given this series of events, we would predict that the tree for the four species should be as shown in Figure 11.4.

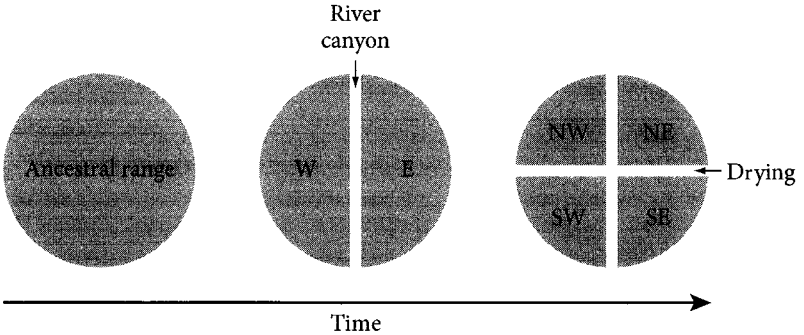


FIGURE 11.3 **A hypothetical example to illustrate vicariance.** An ancestral species' range is divided first by the formation of a north-south river canyon, and later by drying of a strip oriented in the west-east direction.

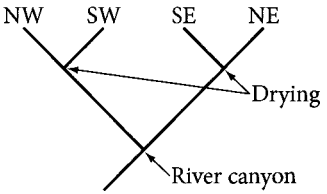


FIGURE 11.4 **The phylogenetic tree expected for the four species arising from the two vicariance events shown in Figure 11.3.** Because the first vicariance event separates the eastern and western species, the east-west divide corresponds to the deepest node in the tree.

Tree topologies that are shaped by vicariance may sometimes corroborate other sources of geological data. For example, geologists have concluded that South America, Antarctica, and Australasia remained attached to one another long after they were separated from other portions of the southern supercontinent Gondwana (Africa, Madagascar, and India), as shown in Figure 11.5.

We see the geologic history of Gondwana mirrored in the phylogeny of ratite birds shown in Figure 11.6. First, the order of branching in the tree follows the order of the separation of the continents, a classic pattern of vicariance. The split between the African taxa and the Australasian/South American taxa occurred first. This is expected if Africa separated from Gondwana while Australasia and South America remained joined. Second, molecular dating (summarized later in this chapter) corroborates the vicariance explanation. A study

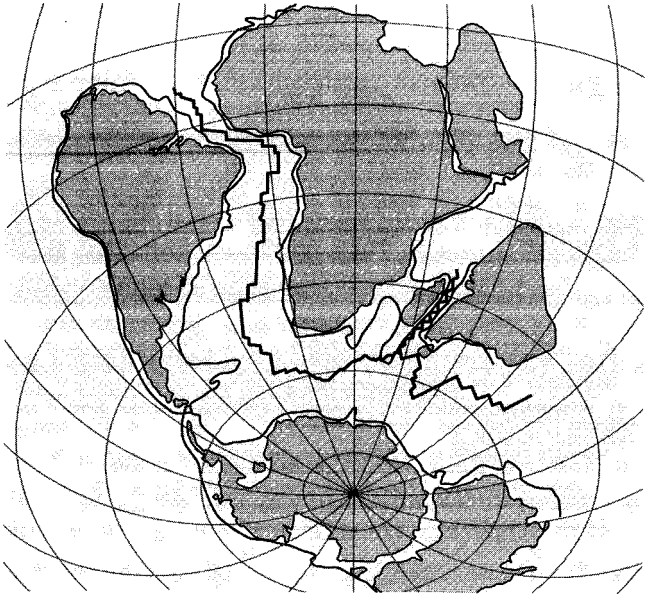


FIGURE 11.5 A reconstruction of the southern continents at ca. 80 Ma as the supercontinent of Gondwana is breaking up. Notice the almost continuous connection between South America, Antarctica, and Australasia. Based on de Wit et al. (1999).

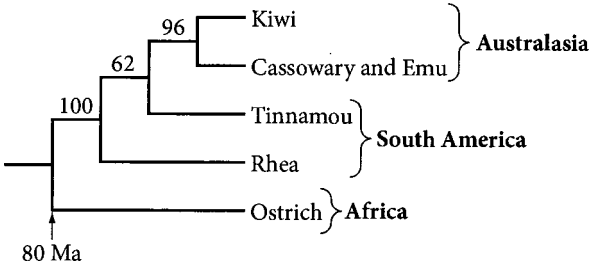


FIGURE 11.6 The phylogeny of ratite birds. Numbers are bootstrap percentages (from Hackett et al. 2008). The node that corresponds, under vicariance, to the separation of the African ratite taxon from those now found in South America, Antarctica, and Australasia is marked. This node has been estimated to be ca. 80 Ma, meaning that it could correspond to vicariance during the Gondwanan breakup (Cooper et al. 2001).

by Cooper et al. (2001) dated these lineages to approximately 80 Ma, about the time that Gondwana was breaking apart.

With these pieces of information, we can envision a scenario in which there was a single ancestral ratite bird that was distributed across all of the southern continents. The separation of Africa caused this ancestral species to split into two lineages: one in Africa and one in South America, Antarctica, and Australasia. Whereas the former has only one living descendant, the ostrich, multiple species persisted in both South America and Australasia (the Moas of New Zealand having relatively recently gone extinct). Although the true history was probably more complex, this hypothetical scenario illustrates the potential for vicariance events to impact the phylogenetic relationships of lineages resident in the affected geographic region.

An extension of the principle of vicariance is that a single geological event, such as continental fragmentation, mountain uplift, river formation, sea-level change, and so on, can simultaneously affect multiple taxa in parallel. This can result in diverse groups of organisms having concordant phylogenetic patterns. For example, we might expect that other groups that radiated around the time that Gondwana fragmented would have a similar phylogenetic pattern to the ratite birds. While still controversial, claims have been made that both marsupial mammals and the Southern Beech tree genus (*Nothofagus*) have phylogenetic trees that mirror the breakup of Gondwana in a manner that is similar to ratite birds.

You may have noticed that in all of the above examples, the history of geographic distribution is inferred using parsimony. This is because model-based approaches to biogeography have only recently been developed. These methods are computationally intensive because of the difficulties of simultaneously modeling geographic migrations, range expansions, local extinction, and vicariance. Nonetheless, now that some model-based methods have become readily available (e.g., in the program Lagrange; Ree et al. 2005), it is likely that future biogeographic research will come to be dominated by these and related statistical approaches.

COEVOLUTION AND COPHYLOGENETICS

Patterns of lineage splitting can also be affected by ecological relationships among species, such as mutualistic or parasitic associations. When one species

is dependent for its survival on a distantly related species, this host species functions rather like a geographic area to which the parasite or mutualist is restricted. As a result, many of the same ideas we developed in relation to biogeography apply to cases of close coevolution.

The geographical problem of dispersal is analogous to the problem of host switching. A lineage of parasites can be expected, once in a while, to acquire a new host. Like long-distance dispersal events, host-switching events are likely to be rare because they too require that founder organisms make it to a new piece of real estate (the body of the new host) and are fortunate enough to be able to survive and reproduce in this novel environment.

Once we have inferred the phylogeny of a group of parasites or mutualists, it becomes possible to reconstruct the history of host switching using methods such as parsimony or maximum likelihood. A simple example is provided by the transition of tapeworms of the genus *Taenia* to humans. A phylogenetic study by Hoberg et al. (2001) used parsimony to reconstruct the origin of the three species of *Taenia* that infect humans. This analysis (Figure 11.7) showed that there were two transitions from carnivoran definitive hosts to humans. Furthermore, by looking at the closest relatives of the human tapeworms, the authors conclude that both events of host switching to humans occurred in Africa and that both occurred soon after humans began eating large, herbivorous ungulates (antelopes, etc.), which serve as the intermediate hosts of these tapeworms.

The coevolutionary situation also has an analog of vicariance. In the same way that geographical vicariance forces a break within a species' range, host speciation tends to limit gene flow between host-adapted races of partner species. This means that lineage splitting within a host can drive corresponding lineage splits in their parasites or mutualists. The result of such ***cospeciation*** is a pair of phylogenetic trees for the two interacting groups that share many nodes in common. This is analogous to the matching of trees found when the evolution of multiple groups has been shaped by the same series of vicariance events. The study of groups that have similar phylogenies due to cospeciation is sometimes called ***cophylogenetics.***

To illustrate the field of cophylogenetics, a nice example is the coevolution of figs and their wasp pollinators. The pollination biology of figs is unusual. Female wasp pollinators enter the fig's enclosed reproductive structure, the syconium, and lay their eggs in a subset of the flowers. After the adult wasps emerge, the flightless males fertilize the females, and only the females (laden

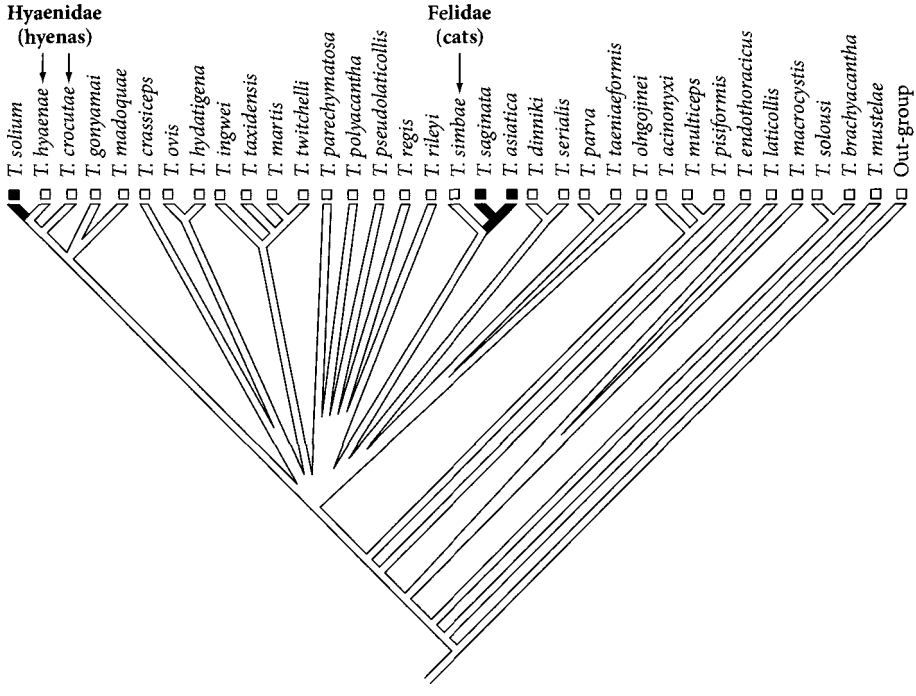


FIGURE 11.7 **Parsimony reconstruction of definitive host on the estimated phylogeny of *Taenia* species.** White branches are inferred to have carnivoran hosts, while black branches correspond to human-infecting lineages. The carnivoran families infected by the tips most closely related to the human parasites are indicated. Adapted from Hoberg et al. (2001).

with pollen) fly off to find another syconium. Because mating happens within the syconium, plant and insect reproduction is closely tied. As a result, it is easy to imagine that speciation of a fig species could result in corresponding speciation of fig wasps. Does this, in fact, apply?

This question has been extensively studied. As an illustration, Figure 11.8 shows the paired phylogenies of a clade of figs and their pollinators. The lines between the tips show which wasp species pollinates which fig species. Apart from the wasp *Ceratosolen appendiculatus*, which appears to have undergone a host-switching event from a relative of *Ficus variegata* to *F. theophrastoides*, there appears to be a general pattern of cophylogeny, especially when one allows for the low bootstrap support for some of the conflicting clades.

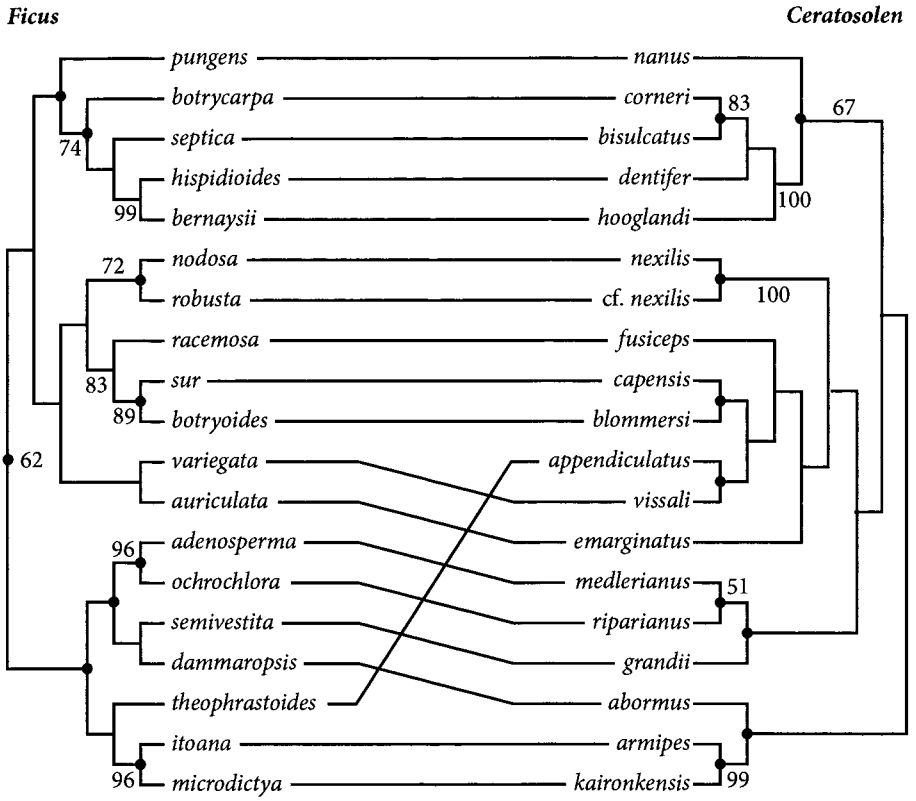


FIGURE 11.8 Estimates of the phylogeny of some fig species (left) and their pollinating wasps (right). The lines between the two trees indicate which wasp species pollinates which fig species. The numbers on the branches are bootstrap percentages. Adapted from Weiblen and Bush (2002).

In many cases, however, it is not visually apparent that the phylogenies of hosts and parasites or the phylogenies of specialized mutualists (e.g., plants and pollinators) match more than would be expected without actual cospeciation. A number of statistical tests have been developed and have been applied to figs/fig wasps and other situations. Even in cases where there have been multiple host-switching events, such analyses may find significantly more cospeciation than you would expect by chance. These results vividly show how interacting species can affect one another's evolutionary trajectory.

PHYLOGENETIC COMMUNITY ECOLOGY

As described in the previous sections, phylogenies are shaped both by geography and ecological interactions. These two considerations come together in the relatively young field of ***phylogenetic community ecology.*** This discipline aims to understand how evolutionary relationships affect the co-occurrence of species in ecological communities. Instead of focusing on pairwise interactions among distantly related species, such as hosts and parasites, phylogenetic community ecology focuses on the role of evolutionary relationships in shaping interactions among large suites of species, which may be closely or distantly related.

Beginning with a seminal study by Campbell Webb (2000), phylogenetic ecologists have been exploring the role of phylogeny in shaping community composition. The general work flow entails inferring the phylogenetic relationships of all species in some major clade that occurs in a particular geographic zone. Then field studies are conducted, which aim to determine which tips of this larger tree occur at particular sites. Finally, the observed distribution of species across communities is compared with the pattern that would be expected if phylogeny had played no role in community composition. Such studies have elucidated two distinct patterns: ***phylogenetic clustering*** and ***phylogenetic overdispersion*** (the latter also referred to as phylogenetic evenness).

Phylogenetic clustering applies when co-occurring species are more closely related to one another than expected by chance (Figure 11.9). This arises when

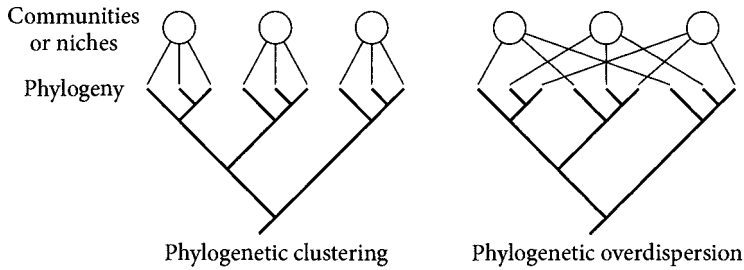


FIGURE 11.9 Phylogenetic clustering versus overdispersion. Clustering occurs when closely related tips tend to be found in the same communities. Overdispersion describes the pattern where taxa are less phylogenetically clustered in communities than expected by chance. Adapted from Cavender-Bares et al. (2004).

members of a particular clade have a tendency to co-occur in the same community. Conversely, overdispersion applies when communities tend to be composed of more distantly related species than expected by chance. This occurs when close relatives are prevented from coexisting, for example, due to competitive exclusion.

A study of plant communities in the fynbos region of South Africa by Procheç et al. (2006) provides an example of phylogenetic clustering (Figure 11.10). The distribution of plant species in 10 × 10-meter plots was analyzed to see if co-occurring species are more or less closely related than expected by chance. For each community Procheç et al. calculated the observed cumulative evolutionary age (CEA): the combined branch length of a community when branches are measured in millions of years. They then compared the observed CEA to the expected CEA under the assumption that communities are assembled randomly. They found that the observed CEAs were significantly lower than expected, meaning that species are more closely related than expected by chance. This pattern is indicative of phylogenetic clustering.

The simplest explanation of phylogenetic clustering is ***phylogenetic niche conservatism,*** which can arise whenever the niche that a species occupies is a heritable trait. In that case, ecology tends to remain conserved through lineage-

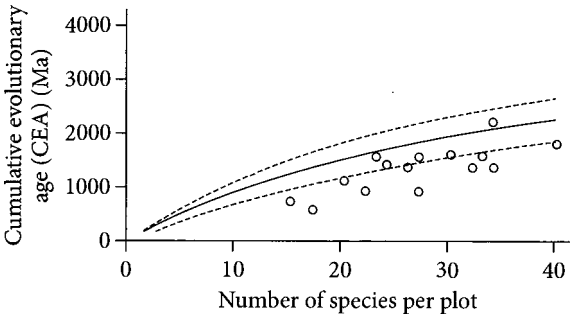


FIGURE 11.10 An empirical example of phylogenetic clustering. The cumulative evolutionary age (CEA) of a community is the sum of the branch lengths represented in the community. Phylogenetic clustering is seen when adding species does not add as much CEA as expected by chance. This graph shows that in fynbos communities in South Africa, plants tend to fall below the 95% confidence interval around the expected CEA (indicated with dashed lines), indicating phylogenetic clustering. Based on Procheç et al. (2006).

splitting events, resulting in clades that are composed of species that occur in similar communities. For example, one clade of plants might be best able to live in xeric habitats with high-salinity soils, while another clade might do best in mesic forests on sandy soils. In the case of plant communities in the fynbos, the low CEA reported by Procheç et al. (2006) is due to overrepresentation of a few clades that seem to have radiated in the fynbos, such as the family Proteaceae.

The converse pattern, phylogenetic overdispersion, is also seen, especially when one looks at very closely related species. For example, Helmus et al. (2007) compared the phylogenetic correlation (i.e., the amount of shared evolutionary history) between pairs of sunfish species with their pattern of occurrence in various pools. After correcting for environmental features of the pools, they found a significant negative relationship between phylogenetic correlation and co-occurrence (Figure 11.11). This shows that more closely related species are less likely to co-occur.

Several alternative explanations of overdispersion have been proposed. It may reflect the fact that closely related species are unable to coexist because they are too similar (e.g., they competitively exclude one another or are susceptible to the same specialized pests or predators). Alternatively, overdispersion

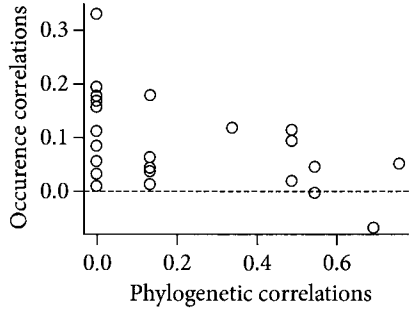


FIGURE 11.11 An example of phylogenetic overdispersion. The horizontal axis shows the phylogenetic correlation of each pair of sunfish, that is, how closely related they are to each other. The vertical axis shows a measure of the frequency of co-occurence of species pairs in the same pool. There is a significant negative association (0.40, $P < 0.05$), which shows that more closely related fish are less likely to co-occur. Adapted from Helmus et al. (2007).

could be explained by allopatric speciation, which tends to result in sister species that are geographically separated.

Overdispersion and clustering tend to be caused by phenomena that occur at different scales. Broad-scale adaptation explains clustering, whereas the lack of coexistence of very closely related species explains overdispersion. As a result, both patterns may be found within a single clade. While there is still much work to be done to understand how the transition between clustering and overdispersion occurs, it is now clear that a full understanding of community composition can benefit by taking phylogenetic history into account.

PRINCIPLES OF MOLECULAR DATING

The nodes of a tree represent actual ancestral populations immediately before they were subdivided, usually by vicariance or dispersal, into two descendant lineages (see Chapter 3). These populations lived at some particular time. If we can infer the actual ages of the nodes, we can also make inferences about when particular traits evolved or when lineages dispersed between landmasses. The field of ***molecular dating*** combines information from the fossil record and analyses of molecular data to obtain a tree with dated nodes.

Molecular data, with very rare exceptions, come from living taxa. This means that all tips are the same number of time units from the root. Likewise, all the tips descended from a particular node have to be the same number of time units away from that node. A tree with all the tips equidistant from the root is called ***ultrametric.*** The challenge of molecular dating is to estimate a chronogram, an ultrametric tree whose branches are proportional to time (Chapter 3). A chronogram may be called a ***relative chronogram*** when the branches are proportional to time but the conversion to actual time is uncertain. A chronogram that has been calibrated to show actual time is an ***absolute chronogram.***

A key feature of all methods that estimate chronograms is that, instead of estimating branch lengths one at a time, they estimate the distance between each node and its descendant tips: the ***nodal depth*** or ***nodal height.*** Although this process usually slows down computation, it has one significant advantage. Because the method identifies the node with the greatest depth, corresponding to the root, it is not necessary to include an outgroup (see Chapter 3). Thus, when we do molecular dating, whether using maximum likelihood or Bayes-

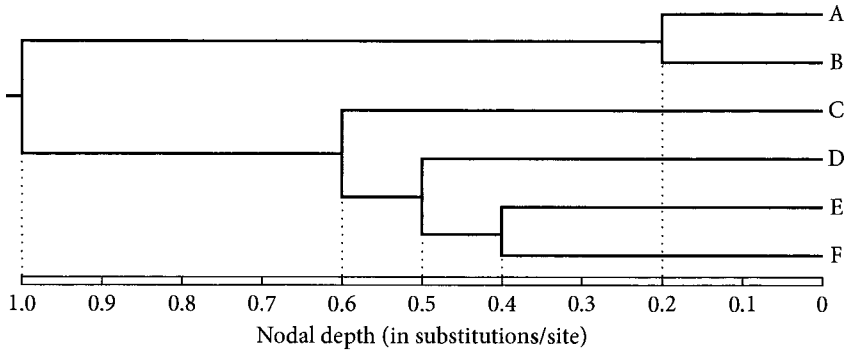


FIGURE 11.12 A relative chronogram showing the depth of all nodes. The depth of the deepest node, the root, is set to 1.0.

ian methods, the output is a rooted tree. A hypothetical example is provided in Figure 11.12.

As discussed in Chapter 8, the natural unit of branch length in a phylogenetic tree is the evolutionary distance, which is the product of the rate of evolution (number of substitutions per site per unit time), denoted μ, and time, denoted t. When we conduct phylogenetic inference using a model-based method, we obtain a tree with branches estimated in units of μt. The challenge of obtaining a chronogram boils down to disentangling μ and t. There are two main methods. We can assume a molecular clock, which means that the rate of molecular evolution, μ, is the same for all branches of the tree. Alternatively, we can use a ***relaxed molecular clock*** (or simply a ***relaxed clock***), which permits μ to vary from branch to branch but does so within some constraints. We will discuss these two strategies in turn.

THE MOLECULAR CLOCK MODEL

Chapter 8 described how we can use maximum likelihood to estimate trees that are optimal (having the highest likelihood) given a specific model of molecular evolution. The models we described dealt with the substitution process, differing in which kinds of substitutions are allowed to vary in rate. However, there is another aspect of a molecular model that we did not explore: branch length.

Molecular clock models differ from their nonclock counterparts in adding the constraint that the tree must be ultrametric.

In a nonclock model we search for a set of branch lengths that together explain the observed molecular data. In a molecular clock model (which can be overlaid on any of the substitution models), each node has a defined distance to its descendant tips: its nodal depth or nodal height. When searching for an optimal tree under a clock model using maximum likelihood, the computer searches for a set of nodal depths that, together, maximize the probability of obtaining the data. Likewise, when using Bayesian MCMC, we obtain a sample of ultrametric trees (relative chronograms) that should approximate the posterior distribution under the clock model.

While clock models have some desirable characteristics, we certainly would not want to use them if they are incompatible with the data. If the rate of evolution has varied across the tree, then application of a clock model could yield misleading date estimates. Fortunately, we can use likelihood ratio tests to decide whether a clock model is justified. The approach is the same as was described in Chapter 8 for substitution models, and in Chapter 10 for deciding between dependent and independent models of discrete trait evolution. As you may recall, the key consideration in such likelihood ratio tests is the difference in the number of free parameters between the competing models. How does this apply to a comparison of clock and nonclock models?

The clock and nonclock models differ in the treatment of branch lengths. In a clock model the free parameters are the nodal depths, of which there are $n-1$ for n tips. In a nonclock model the free parameters are the individual branches, of which there will be $2n-3$. This means that the clock model is actually a simpler model than the nonclock model because it has $n-2$ fewer parameters ($[2n-3]-[n-1] = n-2$). In a manner analogous to how one compares substitution models (e.g., JC versus F81), we can use a chi-square distribution with $n-2$ degrees of freedom to see if the data are significantly better explained by the nonclock model than by the clock model.

As an example of the application of a clock model, Hibbett (2001) analyzed a set of 13 nuclear ribosomal DNA sequences from shiitake mushrooms (*Lentinula*) and their relatives. He found that the optimal tree under the nonclock model had a log-likelihood of −3473.0, whereas the clock model had a log-likelihood of −3479.6. Twice the difference in these log-likelihoods is not judged significant under a chi-square distribution with 11 degrees of freedom.

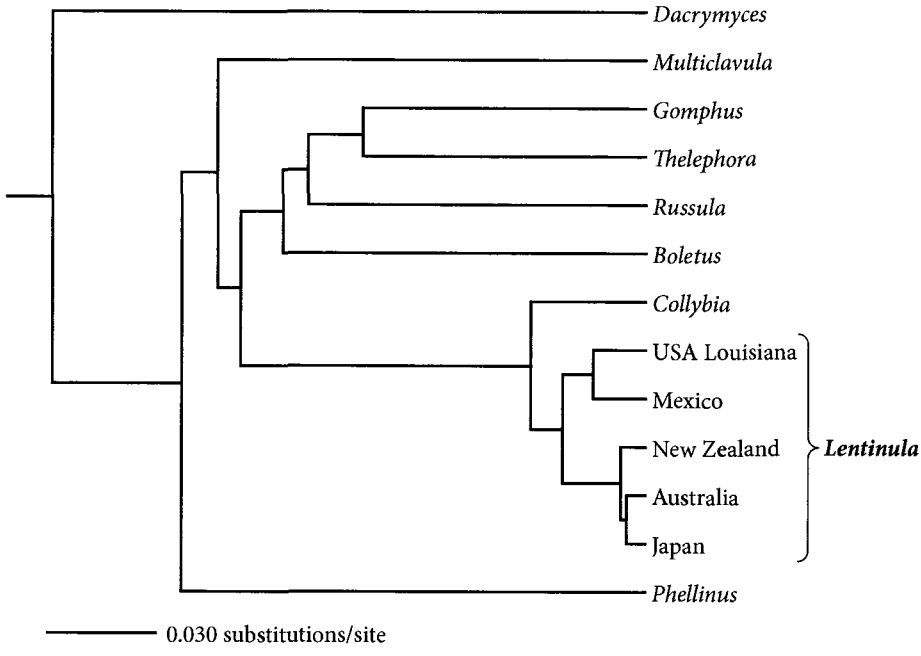


FIGURE 11.13 Relative chronogram for shiitake mushrooms *(Lentinula)* and other Basidiomycete mushrooms obtained under a molecular clock model. Adapted from Hibbett (2001).

This means that the simpler (clock) model is preferable for these data. Figure 11.13 shows the resulting relative chronogram.

For many downstream purposes, it is important to consider uncertainty in the estimates of nodal depth. In a Bayesian framework, no additional steps are required because the posterior sample obtained by MCMC provides an estimate of the credibility interval on each node's depth. In a likelihood framework, the most frequently used way to assess confidence in the relative nodal ages is to use a parametric bootstrap analysis (Chapter 9). This entails evolving data up the optimal clock tree many times. For each pseudoreplicate data set, the nodal depths on the optimal tree topology are estimated by maximum likelihood. By dropping the 2.5% deepest and the 2.5% shallowest depths of each node, a 95% confidence interval on the node's depth can be obtained.

RELAXED CLOCK MODELS

The rate of molecular evolution is influenced by several population genetic factors, potentially including the mutation rate, the strength and form of selection, the recombination rate, and population size. As a result, we should be surprised to find many cases in which the rate of evolution is the same in all lineages of a large phylogenetic tree. Empirically, practitioners have found that the molecular clock model is commonly rejected in favor of nonclock models. For this reason, a set of ***relaxed clock models*** have been developed, which offer a middle ground between clock and nonclock models.

Imagine that we have a tree with branches whose lengths (in units of evolutionary distance, μt) have been estimated. If a molecular clock applied (i.e., if μ is the same for all branches), then the tree would be ultrametric. This would mean that a phylogram madeout of bungee cords could be pinned to a bulletin board such that all the nodes are attached to the board and all the tips are lined up (Figure 11.14). In contrast, if a molecular clock did not apply, the tips would not line up unless we stretched and/or compressed some of the branches. Stretching implies that the rate of evolution was lower on this branch than elsewhere. Conversely, compression implies that the rate of evolution was higher. Relaxed clock methods provide a rational method for deciding which branches to compress or stretch, and by how much.

Over the years, several different methods have been developed, but currently most studies use one of two kinds of relaxed molecular clocks: uncorrelated and correlated. Both permit the rate of evolution to vary from branch to branch but place limits on the range of rates that apply. For computation reasons, these methods are usually implemented in a Bayesian MCMC framework (e.g., using the program BEAST; Drummond and Rambaut 2007). Relaxed clock analysis can be done either on a single, prespecified tree topology or can be integrated into a full Bayesian analysis that also explores the tree topology.

The ***uncorrelated relaxed molecular clock*** model differs from the strict clock model by allowing the rate of evolution to differ on different branches of the tree. However, the relaxed clock model places some constraints on the total amount of rate variation seen across the tree. The variation in the rate of evolution of lineages is drawn from a predefined distribution of rates (for example, a log-normal or exponential distribution). By specifying a prior probability for

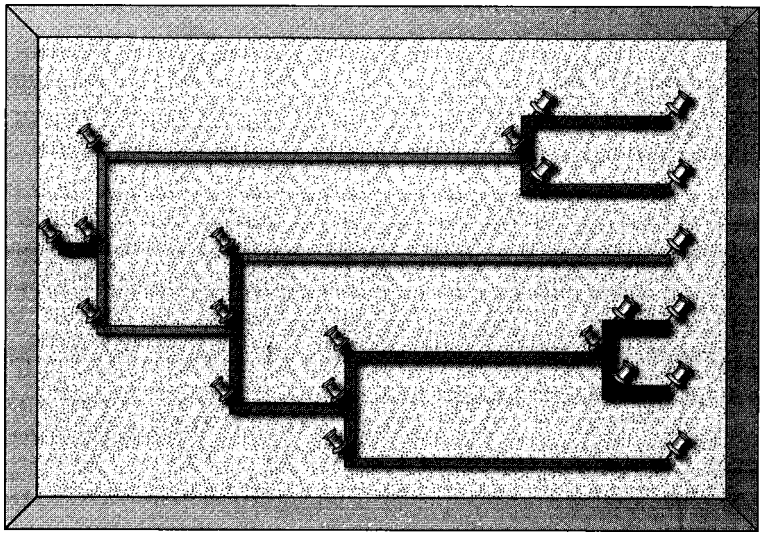


FIGURE 11.14 **Converting a non-ultrametric tree into a relative chronogram.** This process entails stretching some branches (implying that they experienced a slower rate of evolution) and shrinking others (implying a higher rate of evolution). By an appropriate mix of stretching and shrinking, it is possible to adjust any phylogram into an ultrametric form.

the mean and variance of the evolutionary rate, it is possible to use MCMC to obtain an estimate of the posterior probability of each node's relative depth.

The ***correlated*** (or ***autocorrelated) relaxed clock*** model also permits the rate of evolution to vary. However, unlike the uncorrelated model, it assumes that the rate of evolution evolves slowly enough that there is a tendency for branches to share a similar rate of evolution with their immediate ancestral and immediate descendant branches. In effect, the model treats the rate of evolution as a continuous trait evolving across the tree. As with the uncorrelated model, we define a distribution of rates. However, in this case the distribution describes the difference in rate between an ancestral and descendant branch.

Under either of the relaxed clock models, the computer keeps track of both the nodal depths and the rate of evolution of each branch, allowing both to be updated as part of the MCMC analysis. During the likelihood calculations, each branch's length, in units of evolutionary distance, is determined by multiplying its duration and rate of evolution. In this way, the post–burn-in sample from an

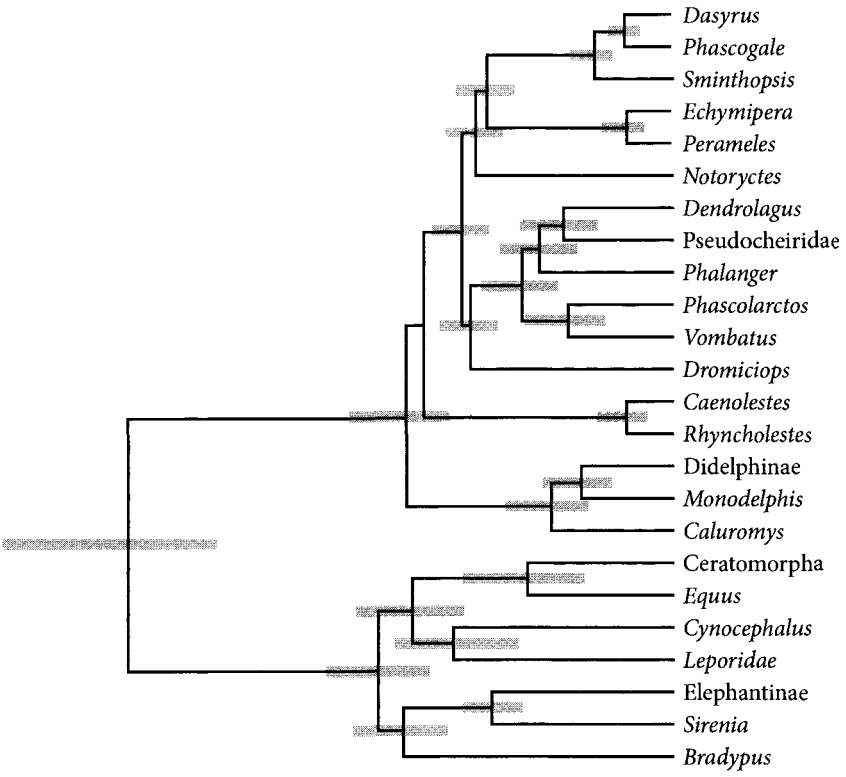


FIGURE 11.15 **A chronogram for marsupial mammals.** 95% credibility intervals around nodal depths are indicated through the use of gray bars. Based on Drummond et al. (2006).

MCMC analysis provides an estimate of the posterior probability distribution of ultrametric trees. From this it is possible to extract an estimate of the 95% credibility interval on nodal depths. Figure 11.15 shows an example of a relative chronogram derived from a relaxed clock analysis (uncorrelated relaxed clock with a log-normal distribution of rates) of marsupials (plus outgroups). The 95% credibility intervals on nodal depths are indicated by bars.

CALIBRATING A CHRONOGRAM

While a relative chronogram has branch lengths in units other than time, they should be proportional to time. This means that if we can determine the actual age of just one node, then the ages of all other nodes can easily be calculated. But how do we determine the age of a node?

One method involves using fossils of known age. As we alluded to in Chapter 3, fossil calibration is not as easy as it sounds. Fossils are unlikely to represent organisms that lived exactly at a lineage-splitting event, but rather they will almost always represent a lineage that diverged from one of the tree's other branches some time before the date of fossilization. Thus, fossils generally only serve to indicate a minimal age for the node that is immediately ancestral to the point of attachment of the fossil (see Figure 3.22).

The problem of calibrating a tree may be further confounded by the fragmentary nature of the fossil record. This is problematic because there can be considerable uncertainty as to where a fossil lineage attaches to the broader tree. Also, there is often some uncertainty as to the actual geological date of a fossil. Thus, single fossils rarely provide convincing nodal ages. Fortunately, we often have multiple fossils, which may help reduce the error rate.

Bayesian molecular dating methods allow us to attach multiple calibration points to a tree. Also, each calibration point can be assigned a distribution of possible ages instead of a single fixed age. In this way, when multiple fossils are included, the age assigned to one calibrated node will be influenced by the ages of all other nodes (and the raw molecular data). Thanks to such methods, and through the use of multiple different genes to study the same branching events, we are increasingly able to obtain reasonably dated absolute chronograms. These, in turn, can be used to test a diversity of interesting biogeographic hypotheses, of which we will present just one of many interesting examples.

Yoder et al. (2003) collected molecular sequence data from two of the four clades of mammals that have diversified on Madagascar. They used Bayesian relaxed clock methods to assess whether these lineages invaded Madagascar across a land bridge from Africa, which might (or might not) have existed 27–45 Ma. Figure 11.16 shows that the Malagasy carnivorans and lemurs migrated to Madagascar at different times and that neither migration event corresponds to the timing of the putative land bridge. This result suggests that, rather than overland migration, the arrival of mammals in Madagascar

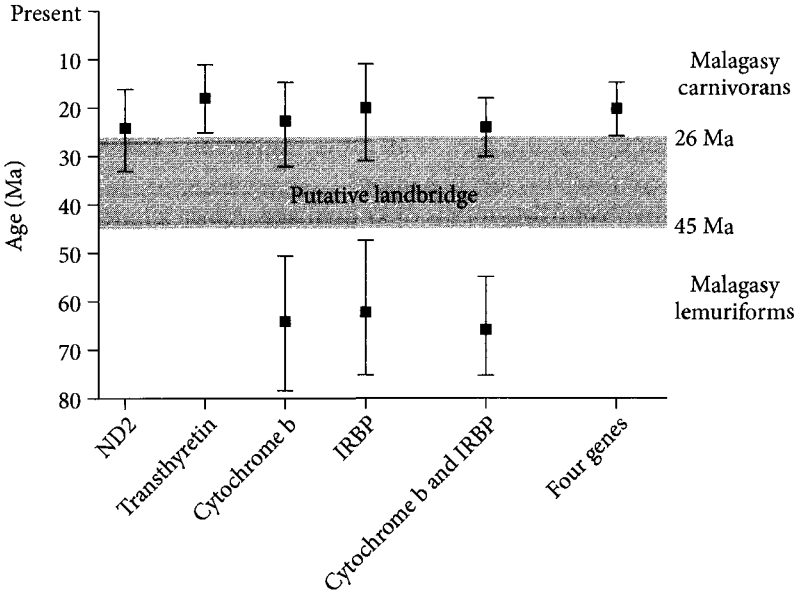


FIGURE 11.16 Dates of the inferred migrations to Madagascar of Malagasy lemuriforms and carnivorans. The squares show the estimated ages from each of one to four genes and their combinations. Error bars indicate uncertainty under the relaxed clock model. Adapted from Yoder et al. (2003).

probably entailed rare over-water dispersal events, for example, via rafting on trees washed down rivers.

CLADE SIZE AND DIVERSIFICATION RATE

Being able to estimate the actual ages of the nodes of a phylogenetic tree is not just useful for reconstructing biogeography, it also helps biologists explore questions involving the rates of lineage splitting and extinction. By knowing the actual ages of nodes, it becomes possible to calculate the rate at which clades have diversified, that is, the average number of extra lineages added to the clade per million years.

Biologists have long been struck by the great variation in species richness across different clades of the tree of life. Why, for example, does the clade Arthropoda (insects, crustaceans, spiders, etc.) contain several million species while its likely sister clade, Onycophora, contains only a couple of hundred species? Does this reflect a higher rate of diversification in Arthropoda than in Onycophora, and if it does, can we pinpoint specific traits of these two taxa that might explain their differential success?

In this section, we will explore how we can quantify diversification rate and how we can identify clades that have been unusually successful, in the sense that they have accumulated more species. In the next section, we will explore how we can identify traits that have increased or decreased the rate of diversification.

Although there has been an increase in the total number of species on the planet since the origin of life, this increase has been very slow. Furthermore, examination of the fossil record suggests that the total number of lineages often remains static for long periods of time, and also that there have been several mass extinctions where diversity has gone down. But, even in periods that have seen constancy in the total number of species, some clades have diversified while other clades have shrunk. Some of these differences could be due to chance. In other cases, clade size differences are surely due to intrinsic differences in rates of diversification. Phylogenetic trees provide essential tools for exploring such clade-to-clade variation in the lineage diversification rate.

The net rate of diversification of a clade is its rate of lineage splitting, or ***lineage birth rate,*** minus its rate of lineage extinction, or ***lineage death rate.*** Clades with positive diversification rates will tend to increase in size, while those with negative rates will tend to decline. If we know the age of a clade and the number of species it contains, we can estimate its diversification rate.

The relationship between clade age and size is simplest to model if we ignore extinction (at least for now) and consider only lineage birth. Under such a pure birth model, there is a fixed lineage birth rate, λ. We will express birth rate as the average number of splitting events expected per lineage per million years.

A complication immediately arises because, whenever there is a lineage-splitting event, an extra lineage is created that is also capable of splitting. As a result, it is simpler to express the model in terms of the expected waiting time between successive splitting events. If we follow a newly formed lineage it will,

on average, split after $1/\lambda$ million years. For the mathematically inclined, the waiting time between lineage-splitting events is exponentially distributed with a mean of $1/\lambda$ (you may note the similarity to the waiting time for nucleotide positions to change state; Chapter 8).

Given this simple, pure birth model, if we start from a single lineage and allow it to diversify, we can predict how many lineages will be present after t million years. The expected number of lineages will be approximately $e^{\lambda t}$. Or, working backward, if we have used molecular dating to determine the age of a clade, and we know the number of lineages now in that clade, we can estimate the diversification rate, λ. Such calculations are, however, confounded by the high variance around the expectation of $e^{\lambda t}$ and by the fact that extinction can significantly alter these expectations.

While we may not be able to estimate λ from clade size without knowing the extinction rate, we can make one reasonable prediction by looking at sister clades. If two sister clades have the same rates of lineage splitting and extinction, then they ought to be the same size. The great difference in the size of the clades Onycophora and Arthropoda thus provides *prima facie* evidence that the latter clade has a higher rate of lineage splitting and/or a lower rate of lineage extinction than the former. While it is true that, without changes in the rate of lineage splitting, sister clades will on average be the same size, there is actually much more variation around this expectation than you might imagine.

If sister clades were the same size, then at every node in a binary tree, the same number of taxa would be attached to each descendant lineage. In other words, the tree would have a high level of ***tree balance,*** such as that shown in Figure 11.17. This contrasts with highly unbalanced trees, wherein sister groups are very different in size, as in Figure 11.18.

Under a pure birth model, we often get very unbalanced trees even when rates of lineage splitting remain constant. The basic reason is that if one lineage is "lucky" and undergoes some early diversification, it has more lineages that are able to diversify later. There is a positive feedback that tends to add lineages to clades that are already large.

Consider an ancestral lineage diversifying under a pure birth model. The first split would generate the two sister clades, whose sizes we are considering. The second split would have an equal chance of occurring on either side of the basal split. It would generate an $n = 2$ clade sister to an $n = 1$ clade. But what

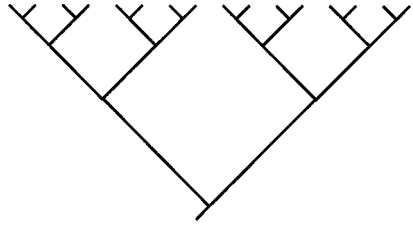


FIGURE 11.17 A maximally balanced tree. At each node, each branch leads to the same number of tips.

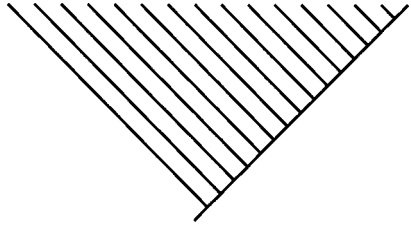


FIGURE 11.18 A maximally unbalanced tree. At all nodes (except one) the number of tips arising from each of the two descendant branches is different.

about the third split? Because it has an equal chance of occurring on each of the three lineages, it has a ⅔ chance of occurring in the $n = 2$ clade and a ⅓ chance of occurring in the $n = 1$ clade. This illustrates that as soon as one clade has, by chance, acquired more lineages than its sister, it will tend to diversify yet further. As a result, unbalanced trees are more likely to occur than intuition might tell you.

We can actually calculate the probability of different tree shapes under the pure birth model. Imagine a clade whose total size is 100 taxa composed of two subclades, A and B. What is the probability that clade A will be size 1 and clade B will be size 99? The probability of no splits happening in clade A is $\frac{1}{2} \times \frac{2}{3} \times \frac{3}{4} \times \frac{4}{5} \times \frac{5}{6} \ldots \frac{99}{100}$. As you can see, the denominator of each term cancels out the numerator of the next term so that this long product simplifies to 1/100. Doing the calculation of the 2:98 split is more complicated, but if you do it you will see that it too comes out to 1/100. So does 3:97, and so on, up to

the 99:1 split. Thus, all possible splits of the large clade into two subclades are equally likely. This shows that even under a constant rate of diversification, a 50:50 split is no more likely than a 1:99 or 99:1 split.

You would not be alone if you find this result counterintuitive. Is it not contradictory to say that diversification rate determines clade size, yet sister clades with the same diversification rate are as likely to be exactly the same size as to be very different in size? However, with some further thought the contradictions can be resolved.

The first thing to note is that, however uneven the two clades' sizes may be, their average should approximate the expected clade size. The high possibility of an unbalanced tree even under a constant rate of diversification does not invalidate the expectation of a clade size of $e^{\lambda t}$; it just shows that there is a large variance in this expectation.

A second point is that we should not focus on the case of an exactly equal number of species in the two sister clades. Rather, we should be asking how frequently very biased splits occur. Under a constant, pure birth model, the probability is only 0.04 that the split between two sister groups will be as biased as or more biased than 2:98 (only 4 of the 100 possible ways of dividing the larger clade have two or fewer species on one side of the basal split). Thus, if we saw a 2:98 split, we would have grounds for suspecting that the rate of diversification might not be constant.

Finally, we should note that looking at tree balance we are focusing only on tree topology and ignoring branch length information. The pure birth model makes specific predictions about the distribution of branch lengths, which affords us much more power to detect unequal rates of diversification than we gain by looking only at tree topology.

The take-home message is that although we can estimate diversification rate from clade sizes and ages, our estimates are often uncertain. Nonetheless, with a big enough chronogram it is possible to estimate the rate of diversification with some confidence and to test the hypothesis that the rate has varied from clade to clade. Through these methods, biologists have begun to identify clades that seem to have unusually high or low diversification rates.

As an example, Alfaro et al. (2009) used an absolute chronogram for the vertebrates and a birth-death model of diversification to identify nine subclades that showed a significantly elevated or reduced rate of diversification. Table 11.1 shows the estimated net rate of diversification (where diversification

TABLE 11.1 Estimated diversification rates of vertebrate clades whose rates of diversification deviate from the background rate of 0.01 lineages/Ma (from Alfaro et al. 2009)

Subclade	Diversification rate (lineages/Ma)
Birds (excepting the ratite and fowl clades)	0.089
Perch and related ray-finned fish	0.082
Placental mammals (excepting the elephant-manatee and sloth-armadillo clades)	0.072
Catfish, carp, and related ray-finned fish	0.053
Lizards and snakes (excepting the gecko clade)	0.040
Euteleost fish (the major group of ray-finned fish)	0.039
Lungfish and coelacanths	0.0017
Tuatara (lizardlike reptiles)	0.0014
Crocodilians	0.00006

equals λ minus the extinction rate) for each of these nine subclades. For comparison, the background rate of diversification for lineages not in these clades was 0.01 lineages/million years. You may notice that none of the major familiar clades (e.g., mammals, birds, lizards/snakes, or ray-finned fish) show an accelerated rate of diversification as a whole. Rather in each case, one specific subclade (or in the ray-finned fish, three subclades) appears to be the source of the larger-than-expected number of species seen today.

KEY INNOVATIONS

The previous section introduced the general problem of detecting whether clades have equal or different rates of diversification. Often, however, we want to go further than this and build a narrative (Chapter 10) to explain *why* certain clades have been so successful in terms of number of species.

An elevated rate of diversification could result from the invasion of a region or landmass that permits the accumulation of species. For example, the new region may contain many empty ecological niches into which the lineage can radiate, or it may have physical features that promote species accumulation (e.g., by being highly fragmented). Alternatively, or in addition, diversification can be spurred by the evolution of particular traits that influence the probability of lineage splitting and/or extinction. This phenomenon, ***character state–dependent diversification,*** arises when a derived character state that goes to fixation in a lineage results in that lineage being more or less prone to undergo lineage splitting or extinction in the future. A trait that accelerates lineage diversification is called a ***key innovation.*** This term is also used to describe traits that allow a lineage to enter a new adaptive zone (e.g., a transition from water onto land), but we will focus on the phylogenetic meaning of key innovation.

A key innovation can act by either increasing the lineage-splitting rate or decreasing the extinction rate. A key innovation can increase the lineage-splitting (speciation) rate by facilitating allopatric isolation of populations (e.g., by restricting gene flow yet still allowing occasional long-distance dispersal), by increasing the probability that populations will evolve reproductive barriers, or by increasing the capacity of a lineage to adapt to diverse ecological niches (reducing interspecific competition). Alternatively, a key innovation could make a lineage less prone to extinction, for example, by increasing population size, adapting organisms to an especially stable environment, or equipping organisms to cope (physiologically or evolutionarily) with changing environments. Since many traits will act in one of these ways, there can be no doubt that key innovations exist. But how can we discover them?

The study of key innovations is similar to the study of adaptation, introduced in Chapter 10. This is not surprising given that key innovations can be considered adaptations at the level of lineages: they promote lineage fitness in much the same way that conventional adaptations promote organismic fitness. In parallel to the case of adaptation, there are three steps in testing a key innovation hypothesis.

First, we use experimental and theoretical analyses to justify the conclusion that the trait really should enhance net diversification rate (lineage fitness). Second, we look to see if the trait arose in a context that allowed it to have this effect, which amounts to showing that the clade with the key innovation has an

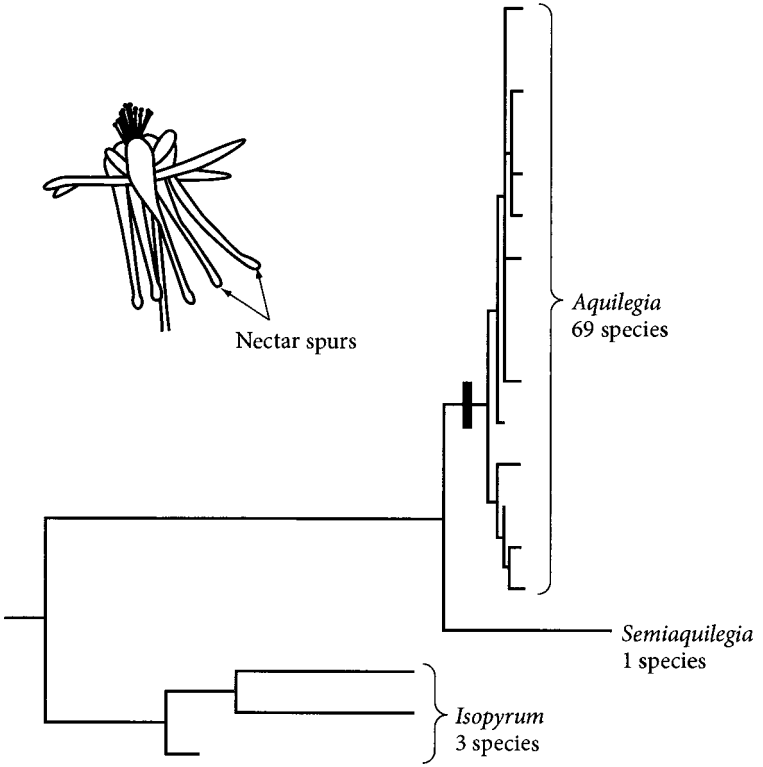


FIGURE 11.19 **A phylogeny of columbines *(Aquilegia)* and outgroups *(Semiaquilegia* and *Isopyrum).*** The black bar marks the inferred origin of nectar spurs. The short branches within *Aquilegia* suggest a rapid radiation, consistent with the hypothesis that nectar spurs are a key innovation. (*Inset*) A flower with nectar spurs. Adapted from Hodges (1997).

elevated rate of lineage diversification compared to its closest relatives without the key innovation. Third, we look for similar but nonhomologous versions of the traits to see if there is a general correlation with elevated diversification rate.

One of the best-studied examples of a key innovation is the nectar spur (a long, narrow outgrowth of a petal containing nectar) of columbine flowers (Figure 11.19). Evolutionary changes in the shape, size, or orientation of nectar spurs can result in rapid shifts in pollinators. This, in turn, could be expected to facilitate lineage splitting by preventing recently separated lineages from

secondarily fusing. Thus, theoretical considerations predict that nectar spurs might increase the diversification rate of columbines.

Hodges and Arnold (1995) tested this hypothesis by assessing whether columbines (the genus *Aquilegia*) do indeed have a higher rate of diversification than the background rate for the spurless clade in which they are embedded. Figure 11.19 shows a phylogram for some columbines and their nearest relatives. The number of species in each clade is shown to the right. With so many more species in *Aquilegia* than in its sister group, *Semiaquilegia*, you will not be surprised that the data support an elevated rate of diversification in the spurred clade. Ree (2005) reached the same conclusion by noting the much shorter branch lengths in *Aquilegia* than in the broader clade.

Even if we accept that columbines have a higher rate of diversification than their spurless relatives, this does not show that the nectar spur is the cause. Perhaps there is some other trait that is responsible for accelerated diversification in columbines. One way to address this possibility is to see if other clades that have flowers with nectar spurs are also more species-rich than their spurless sister clades.

Hodges (1997) documented seven other origins of spurs in flowering plants. In all but one case, the spurred clade contains many more species than its sister clade. This association between nectar spurs and large clades was too frequent to be attributed to chance, suggesting that, indeed, these structures serve as key innovations. This illustrates that phylogenetic analysis not only sheds light on why certain traits evolved but also helps explain why some clades contain more species than others.

JOINT ESTIMATION OF CHARACTER EVOLUTION AND DIVERSIFICATION

The traditional methods for assessing key innovation hypotheses, introduced in the previous section, assume that the history of character state change is known. Conditional on that history, these methods can compare the number of species in lineages with and without a certain character state to test its effect on diversification rate. One major flaw with this approach arises if a character state has indeed affected the shape of the tree. The methods that are conventionally used to infer ancestral states, parsimony or maximum likelihood (as described

in Chapter 10), implicitly assume that traits do not affect lineage diversification. This means that we run the risk of using an incorrect understanding of trait evolution when we test for character state–dependent diversification—specifically in those cases where the trait *does* affect diversification. This biases the traditional tests of key innovations and means that the results of such tests should be interpreted with caution.

Statistical tests have been developed recently that overcome this bias in the study of state-dependent diversification (Maddison et al. 2007). These methods jointly model trait evolution (allowing for different rates of evolution in the two directions) and state-dependent effects on the rate of extinction or lineage splitting. The models are computationally challenging because, in order to model extinction, it is necessary to consider lineages that existed but are not known because all their descendants went extinct before the present. A further challenge arises because unequal rates of gain and loss of a character state (an evolutionary trend) and state-dependent effects on diversification leave rather similar signatures on the data. For example, the observation of a higher number of taxa in state 1 relative to state 0 could arise by differences in transition rates or diversification rates. If the rate of gaining state 1 is much higher than the rate of losing it, the number of taxa in state 1 will come to outnumber those in state 0. Similarly, if the diversification rate is higher in state 1 than in state 0, the result will be a higher proportion of tips in state 1.

The details are not critical, but the simplest joint diversification and transition rate method uses the ***binary-state speciation and extinction (BiSSE) model,*** which is summarized in Figure 11.20. As you can see, it considers one character with two possible states and includes two transformation rates as well as distinct extinction and speciation rates for each character state. Because the temporal pattern of events is critical to disentangle trait evolution and diversification,

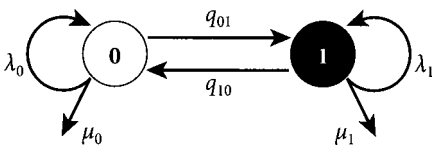


FIGURE 11.20 The binary-state speciation and extinction model, BiSSE. This model allows for different rates of transition between the two states (q_{01} and q_{10}) and also allows that the two states could have different rates of diversification (λ_0, λ_1) and extinction (μ_0, μ_1).

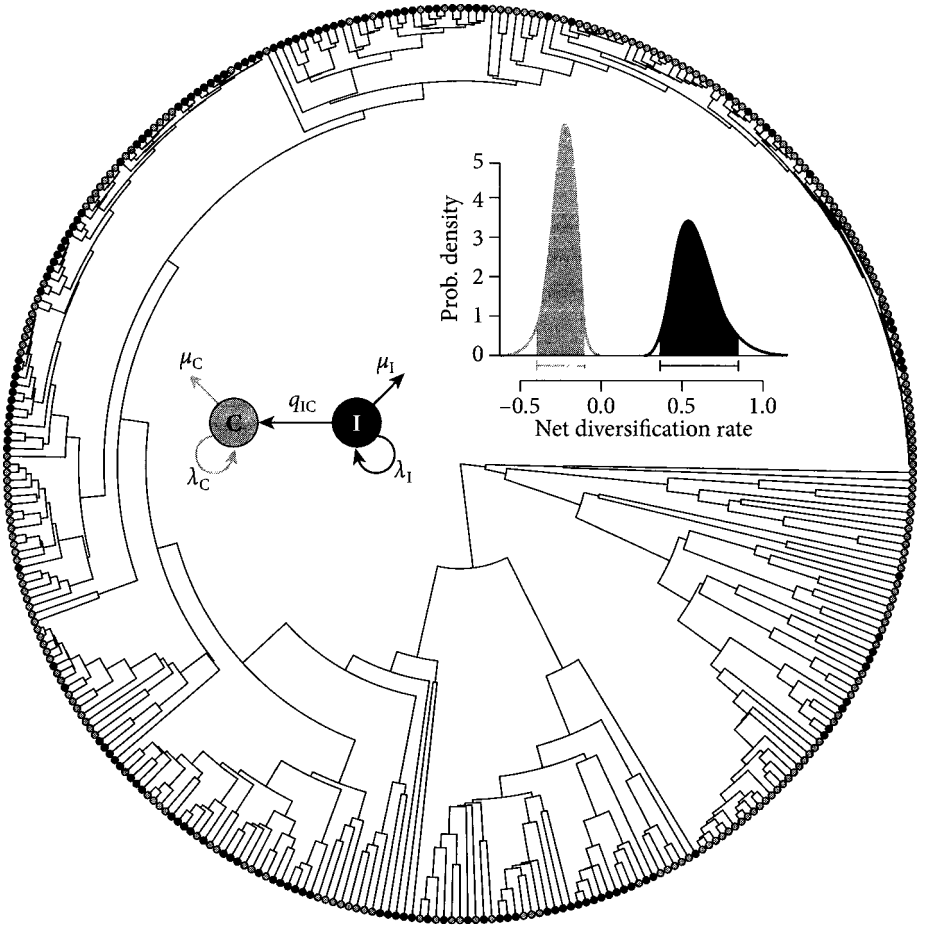


FIGURE 11.21 BiSSE analysis of self-incompatibility (I) and self-compatibilty (C) in Solanaceae. The phylogeny shows the distribution of C (gray) and I (black) species. The posterior distributions for the net diversification rate of C lineages (gray curve) and I lineages (black curve) are inset. Adapted from Goldberg et al. (2010).

BiSSE analyses must be conducted with ultrametric trees (relative chronograms). Furthermore, because of the many parameters and the subtlety of the signal being extracted from the data, trees with many scored taxa are typically needed to obtain convincing conclusions.

An excellent example of the application of BiSSE is provided by Goldberg et al. (2010). Looking at self-compatibility in the tomato family, Solanaceae, they used BiSSE to simultaneously estimate the rate of transitions to self-compatibility (C) from self-incompatibility (I) and the diversification rates of C and I lineages. It was already known that the rate of evolving I from C is much lower than the reverse—in fact, within Solanaceae the rate at which I evolves from C is zero. With this model they found that C lineages diversify at a significantly lower rate than I lineages (Figure 11.21). This result helps to explain the persistence of I lineages over time. If lineages could only go from I to C (because the reverse rate is zero), all lineages would eventually become C. However, because C lineages diversify more slowly than I, both states can be maintained over time.

FURTHER READING

Biogeography: Platnick and Nelson 1978; Linder and Crisp 1995; Morrone and Crisci 1995; Ronquist 1997; Humphries and Parenti 1999

Community phylogenetics: Webb et al. 2002; Ackerly 2004; Cavender-Bares et al. 2004, 2009

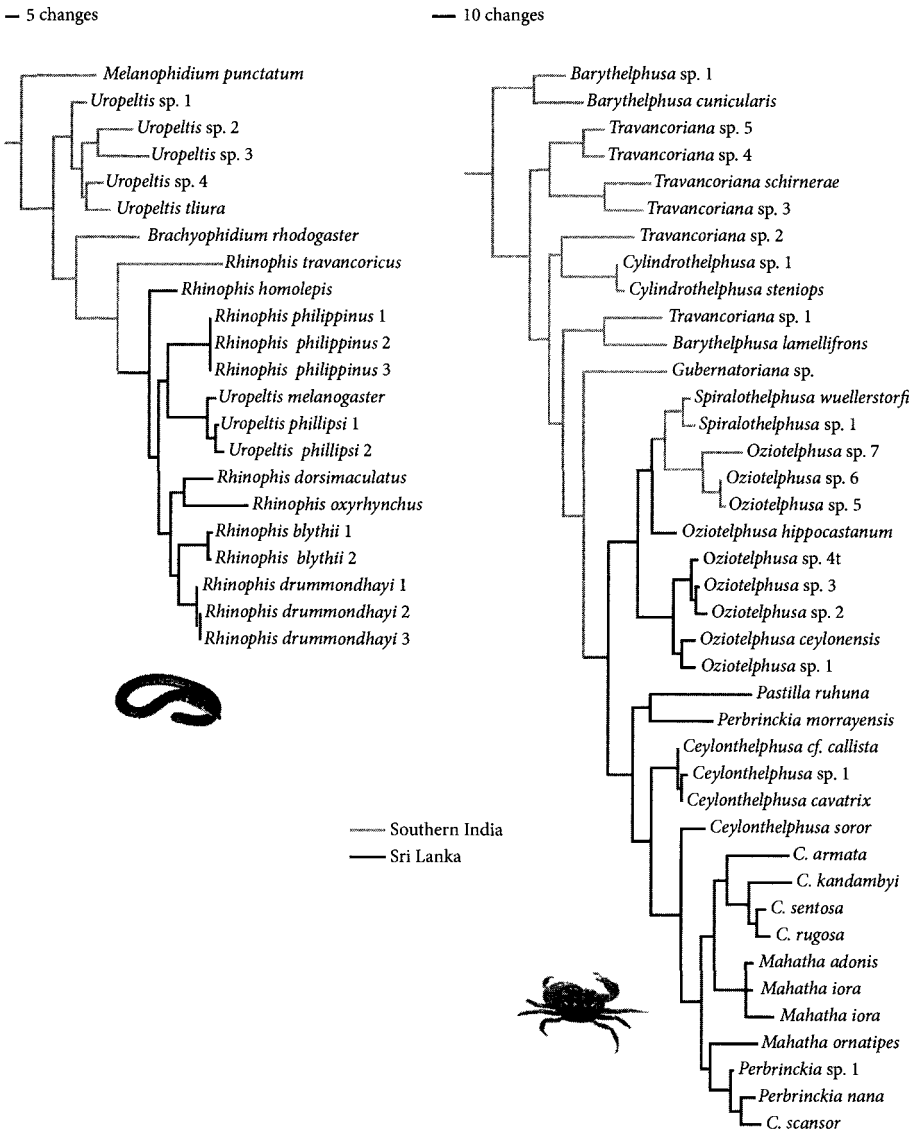
Coevolution and cophylogenetics: Brooks 1981; Huelsenbeck et al. 1997; Page 2003

Molecular dating: Thorne et al. 1998; Sanderson and Doyle 2001; Sanderson 2002; Drummond et al. 2006

Diversification: Slowinski and Guyer 1993; Sanderson and Donoghue 1994; Maddison et al. 2007; FitzJohn et al. 2009; Schwander and Crespi 2009

CHAPTER 11 QUIZ

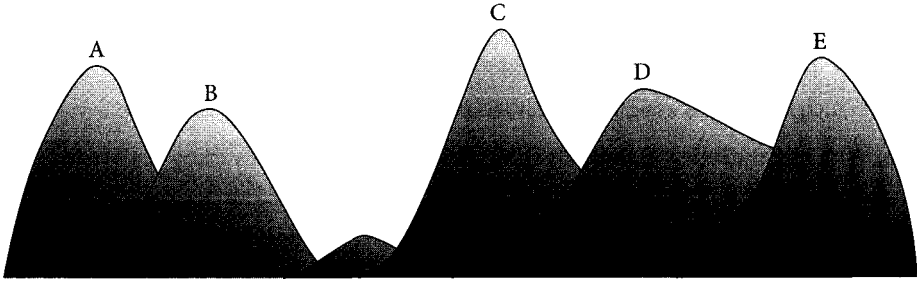
Questions 1–2. The trees below, from Bossuyt et al. (2004), show the phylogeny for two groups of animals (Uroptelid snakes and land crabs) and the distribution of lineages in southern India and Sri Lanka mapped onto the tree with parsimony.

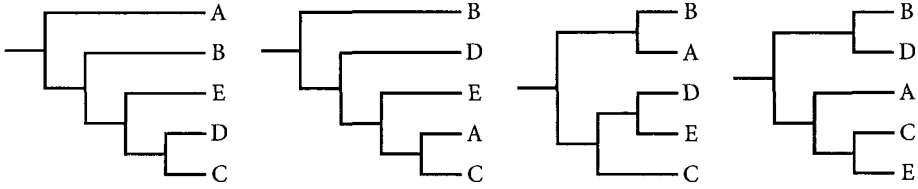


1. Combining these two trees, how many times did a lineage disperse from India to Sri Lanka?
   a. 0 b. 1 c. 2 d. 3 e. The number cannot be determined from these data.

2. What do the two trees say about whether Sri Lanka and southern India were formerly connected and whether the current distributions of species shown might have resulted from the splitting of Sri Lanka from India?
   a. The fact that many southern Indian lineages/clades lack sister groups in Sri Lanka tends to argue against the vicariance hypothesis.
   b. The vicariance hypothesis is contradicted by the fact that the snakes and crabs are of a similar age.
   c. Because both trees have the same pattern of a major Sri Lankan clade emerging from a southern Indian ancestor, the vicariance hypothesis is supported.
   d. There are three cases in which Sri Lankan clades are sister to southern Indian clades, which supports the vicariance hypothesis.
   e. The patterns are not relevant to the hypothesis of vicariance.

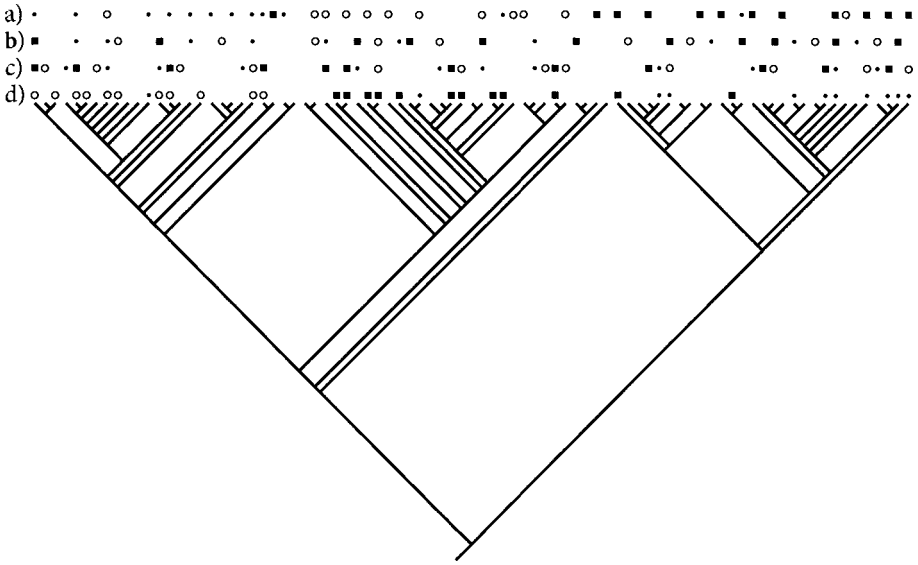
3. The plot below shows the cross section of a hypothetical mountain chain. Five species of gentian are restricted to the five peaks, labeled A–E. Assume that these five species descended from a single widespread ancestral species whose range was fragmented as global warming gradually increased the lowest altitude at which gentians can grow. Given the implied vicariant history, what tree would we predict for these five species?





(a) (b) (c) (d)

Questions 4–6. The hypothetical tree below contains 85 species as tips. The presence of these species in three communities is marked by symbols (●, ■, and ○) aligned with the tips. Four possible assignments of species to communities are shown.

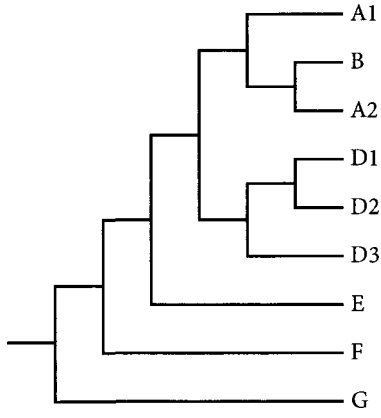


4. Which of the four assignments shows the clearest signal of phylogenetic clustering only?

5. Which of the four assignments shows phylogenetic overdispersion only?

6. Which of the four assignments shows both phylogenetic clustering at a broad taxonomic scale and overdispersion at a lower taxonomic scale?

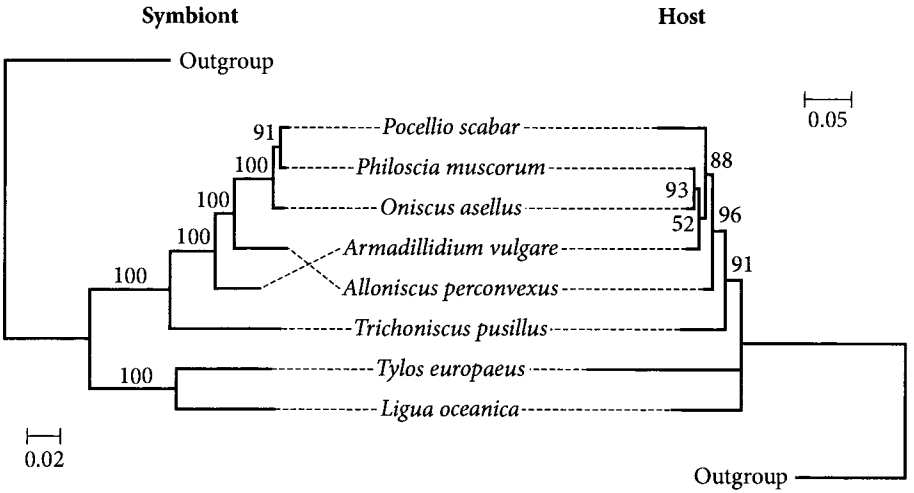
7. The tree below summarizes the phylogeny inferred for the malaria parasite, *Plasmodium*, and its close relatives. The data show the definitive host (host) and secondary host (vector) for each tip in the tree (from Martinsen et al. 2008). Using equally weighted parsimony, how many host-switching events are implied among definitive hosts and vectors?



Tip	Host	Vector
A1	Mammals	Mosquitoes
B	Mammals	Mosquitoes
A2	Mammals	Biting midges
D1	Lizards	Mosquitoes
D2	Birds	Mosquitoes
D3	Birds	Mosquitoes
E	Birds	Biting midges
F	Birds	Louse flies
G	Birds	Black flies

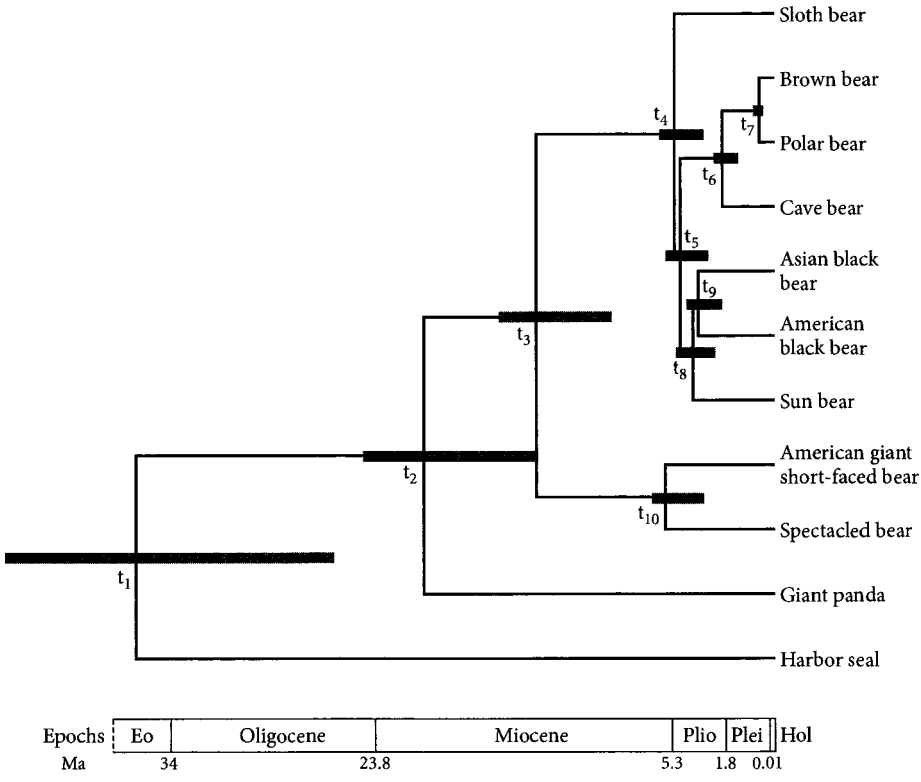
a. One switch of definitive host, two switches of vector.
b. Two switches of definitive host, three switches of vector.
c. Two switches of definitive host, four switches of vector.
d. Three switches of definitive host, four switches of vector.
e. The number of changes cannot be determined by parsimony.

8. The figure (adapted from Fraune and Zimmer 2008) shows the phylogeny for some isopods (crustaceans related to woodlice) and their bacterial endosymbionts. The numbers on the branches are Bayesian clade credibilities, in percent. Allowing that any clade with a clade credibility of less than 95% can be collapsed, what is the *minimum* number of host-switching events that need to be invoked to reconcile the host and symbiont trees?
   a. 0 b. 1 c. 2 d. 3 e. 4



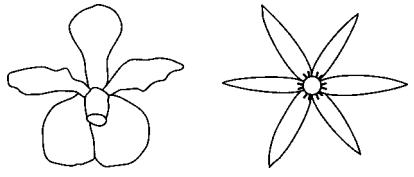
Questions 9–10. The chronogram on the facing page (adapted from Krause et al. 2008) shows the timing of diversification of bears. For each node, the estimated mean age and 95% credibility interval is given.

9. Given this figure, which of the following can be stated with confidence?
   a. The sloth bear is older than the spectacled bear.
   b. The last common ancestor of the American and Asian black bears lived after the last common ancestor of brown and polar bears.
   c. The last common ancestor of the spectacled and the American giant short-faced bears could have lived at the same time as the last common ancestor of the sloth and sun bears.
   d. The common ancestor of the sun bear and the giant panda lived at least 19.09 Ma.
   e. The common ancestor of the cave bear was not present during the Miocene.



10. In the bear phylogeny, imagine we found a new fossil bear, *Ursus cyaneus*, dated at 15 Ma. Based on its morphology we confidently place the fossil as the sister species to the sloth bear. What would this suggest about the dating shown in the figure?
    a. The node "t_4" is probably much too young and should at least be 15 Ma.
    b. The node "t_4" is accurately dated with the sloth bear arising around 5.39 Ma, after the split with *Ursus cyaneus*.
    c. The node "t_5" is probably much younger than 5.05 Ma.
    d. The common ancestor of the sloth bear and the giant panda is at least 30 Ma.

11. Flower symmetry is a trait hypothesized to increase speciation rates because bilaterally symmetrical (zygomorphic) flowers (like the orchid on the left below) have more precise pollen placement than radially symmetrical (actinomorphic) flowers (like the orchid sister group on the right). Sargent (2004) surveyed the number of species in 19 zygomorphic clades and their actinomorphic sister groups. In 15 of the 19 cases, the zygomorphic clade included more species than the actinomorphic group. The probability of a fair coin landing on 15 or more heads in 19 tosses is approximately 0.001. Given this, what might we conclude about the hypothesis?
    a. Actinomorphic taxa probably have a lower extinction rate.
    b. Zygomorphic flowers evolve significantly more commonly than actinomorphic flowers.
    c. Flower symmetry has a higher rate of evolution than expected for a random trait.



    d. If flower symmetry had no effect on diversification rate, we would not expect so many zygomorphic clades to be larger than their sister groups.
    e. The hypothesis is rejected because there is very little chance that differences in speciation rate could explain the observed data.

12. Sexual reproduction is a trait that varies across the tree of life. Many organisms are unable to reproduce without sex, but others are capable of reproducing asexually. Some evolutionary biologists have suggested that the loss of sex (the transition to a purely asexual mode of reproduction) is an evolutionary dead end because such lineages would not have the ability to recombine their genes to create new and advantageous combinations. How could we use the key innovation test to answer this question?

13. The BiSSE model makes it possible to understand how transition rates, speciation rates, and extinction rates together shape the distribution of character states at the tips of the tree. Imagine you were interested in a particular trait, seed dispersal, and the effect of two states (animal and wind dispersal) on diversification. As an exploratory analysis, you conduct a simulation using the BiSSE model with the following parameters (where animal dispersal is 0 and wind is 1): $\mu_0 = \mu_1 = 0$, $\lambda_0 = 0.1$, $\lambda_1 = 0.2$, $q_{01} = 0.05$, $q_{10} = 0.03$. What pattern of distribution of traits among tips would you expect to find?

APPENDIX ONE

Data for Phylogenetic Analysis

The success of phylogenetic inference depends in large measure on the choice of trait data, the accuracy of those data, and the quantity of data obtained. Whether or not you plan on being engaged in phylogenetic analysis, it is important to know the kinds of data that are typically used in phylogenetic analysis, and to understand how data are organized to permit phylogenetic analysis.

Information for phylogenetic inference can come from any kind of trait that is believed to evolve within the constraints of the underlying phylogenetic tree. A trait that varies among tips and shows some degree of heritability (ancestors tending to resemble descendants) has the potential to provide phylogenetic information.

Prior to modern molecular methods, phylogenetic analysis was conducted primarily on morphological variation. Nowadays, the great majority of phylogenetic analyses are conducted based on DNA sequences obtained from representative individuals.

Although this book has focused on DNA sequence and morphological data, there are many other kinds of data that can be used for phylogenetic analysis. In most cases these other kinds of data can be analyzed using methods developed for DNA sequences or morphology, but in some cases specialized methods need to be used. It is beyond the scope of this book to explore all of these methods. However, below we review most of the kinds of data that are used for phylogenetic inference and briefly discuss how they are typically analyzed.

MOLECULAR SEQUENCES

Sequences of peptides or small proteins were the first kind of molecular sequences to be used for phylogenetics. Protein sequences are still used for phylogeny reconstruction, especially to study ancient relationships. This is because

protein sequences are often subject to functional constraints that slow down the rate of evolution. The general principles of phylogenetic analysis of protein sequences are very similar to those applied to DNA sequences.

In the same way that transitions occur at a higher rate than transversions, some amino acid substitutions are expected to occur at a higher rate than others. There are two main reasons for this. First, some amino acid substitutions can be accomplished by changing a single nucleotide in the encoding DNA sequence (e.g., methionine [ATG] to arginine [AGG]), whereas other amino acid substitutions require as many as three nucleotide substitutions (e.g., methionine [ATG] to tyrosine [TAT or TAC]). Additionally, some amino acids have similar physical and chemical properties, so these substitutions are less likely to be weeded out by selection. Biochemists have developed tables that summarize the relative frequency of different kinds of amino acid substitutions for different kinds of proteins. These substitution matrices are sometimes used to guide sequence alignment and to improve phylogenetic analysis of amino acid sequences (e.g., using generalized parsimony, or amino acid maximum likelihood models).

The second kind of molecular sequences to be collected widely were ribosomal RNA (rRNA) sequences. This was because ribosomes could be purified and their RNA could be sequenced with reverse transcriptase. Nowadays, rRNA or messenger RNA (mRNA) sequences are more commonly obtained by sequencing the encoding DNA (or in the case of mRNA, by copying it back into DNA).

Because there is usually a one-to-one correspondence between RNA and DNA sequences, phylogenetic analysis of RNA sequences is basically identical to that of DNA sequences. The only complication that arises is that some RNA molecules adopt folded structures due to bonding between bases at different positions in the molecule. Such interaction between different positions in the sequence is problematic because it violates the general assumption in phylogenetics that each base constitutes an independent piece of historical information. Methods, beyond the scope of this book, have been developed to analyze the coevolution of such paired bases.

All kinds of DNA sequences are used for phylogenetic analysis. In prokaryotes there is generally only a single genome, whereas most eukaryotes have at least one additional organellar genome (mitochondrial or plastid). For historical and technical reasons, animal phylogenetics has focused extensively on

mitochondrial DNA (mtDNA) and plant phylogenetics has often focused on plastid DNA (pDNA, chloroplast DNA, or cpDNA). These organellar genomes are easy to work with because they are present in many copies in each cell, making their isolation relatively easy. Most regions of mtDNA evolve rapidly in animals, meaning that mtDNA has proved informative for studies within species or among groups of closely related species. Likewise, pDNA has come to be widely used for studying closely related plant species because (unlike plant mtDNA), the plastid genome has several rapidly evolving noncoding spacer regions that can be easily sequenced. For examining deeper relationships, phylogeneticists primarily use coding regions of nuclear, mitochondrial, or plastid DNA.

MOLECULAR PRESENCE/ABSENCE DATA

Over the years a number of different molecular techniques have been developed that allow researchers to score organisms for the presence/absence of particular molecular markers. The first two methods, restriction fragment length polymorphisms (RFLPs) and isozymes, date back to a time when molecular sequencing was not feasible. Isozyme data, based on differences in the mobility of proteins in gels, is optimized for population genetic studies and is not well suited to phylogenetic analysis. In contrast, RFLP data, based on differences in restriction sites between sequences, can provide robust phylogenetic information.

The heyday of RFLP analysis for phylogenetics was the 1980s. Methods were developed that allowed scientists to determine the sizes of fragments into which DNA is cut by restriction enzymes (enzymes that cut DNA only if it has a target sequence, typically of 4–6 bases in length). Phylogenetic analysis could be conducted on the presence/absence of fragments of a particular size or, more rigorously, by mapping the actual location of each restriction site and scoring species for the presence or absence of the restriction site in a particular position. These methods were quite effective. However, for all of its historical importance, RFLP analysis was superseded by the development of inexpensive and efficient DNA sequencing methods.

RFLP data can be analyzed as presence/absence data using methods appropriate to morphological characters. However, in so doing, efforts should be

made to take account of an inherent asymmetry in RFLP evolution: shared restriction sites are more likely to be lost in parallel than gained in parallel. This is because an entire restriction site (typically six base pairs long) must match for an enzyme to cut, but any one mismatch will result in a failure to cut. As discussed in Chapter 7, parsimony can be modified using character state weighting (with gains counting more than losses) to partially account for this phenomenon. Alternatively, mathematical models of RFLP evolution have been developed that allow for analysis using maximum likelihood.

A number of distributed molecular markers, which were originally developed for studying variation within populations, have come to be used, at least occasionally, for phylogenetic inference. These include randomly amplified polymorphic DNA (RAPDs), amplified fragment length polymorphisms (AFLPs), and microsatellites. Distributed molecular marker methods quickly scan a set of taxa for molecular variation distributed anywhere in the whole genome. These markers (of which AFLP is currently the most widely used) have been employed for phylogenetic studies of closely related organisms, where it can be difficult to find sufficiently variable gene regions to sequence.

In addition to technical problems related to the identification of homoplasy, distributed molecular markers have one significant drawback for studying phylogenies among closely related species. As discussed in Chapter 6, phenomena such as incomplete lineage sorting mean that, when studying very closely related organisms, different parts of the genome can have different phylogenetic trees. The problem with presence/absence data is that there are not (yet) species tree methods (Chapter 9) available. As a result, while distributed molecular data can yield a resolved tree even at a very low taxonomic scale, it can be uncertain how much of the genome has tracked the resulting tree. Therefore, users of distributed marker data need to be extra vigilant to avoid reading too much significance into the trees they obtain in cases where discordance among gene trees is suspected.

MOLECULAR STRUCTURAL DATA

Whereas molecular marker methods have generally been motivated by the search for variation at low taxonomic scales, structural molecular characters have mainly been of interest because they tend to evolve very slowly. A rare

structural molecular feature shared by a group of taxa can provide compelling evidence that those taxa form a clade.

Structural molecular characters include insertions and deletions, inversions, and duplicative or nonduplicative translocations. For each kind of structural mutation, there is a range of scales from very local mutations (e.g., single base-pair deletions; 6-base-pair inversions) to large-scale ones (e.g., deletion of a whole gene; insertion of an intron; translocation of an entire chromosomal arm). As a general rule, finding that two taxa share a small/local structural mutation is considered weaker evidence of a close relationship than sharing a larger structural characteristic. This is because the probability of homoplasy is higher in the former case. It is more likely that two taxa independently underwent a deletion of the same pair of bases than they independently experienced an insertion of the same 500-base-pair sequence at the same point in the genome.

Structural characters are usually identified by gene or genome sequencing, restriction mapping, or (now, rarely) microscopic observation of chromosomes. In some cases the kind of structural mutation involved can be unambiguously inferred. For example, when a large region of otherwise quite similar sequence is inverted in some taxa relative to others, a molecular inversion is clearly implied. In other cases a diversity of different mutational processes could have contributed to an observed pattern. For example, if taxa differ in their gene order along a chromosome, different combinations of duplications, translocations, inversions, and deletions might provide competing explanations.

The use of molecular structural characters for phylogenetic inference generally follows one of two approaches. The first involves scoring tips for the presence or absence of a number of structural characters and then using parsimony or other standard phylogenetic methods to infer the tree that best explains the full set of characters. This approach allows that any of the characters could have been subject to homoplasy.

The alternative approach is invoked when a particular structural character is considered to have resulted from such an improbable event that homoplasy is ruled out. In that case the Hennigian logic can be invoked (Chapter 7). This means that once the structural character is polarized (e.g., by looking in outgroups), a clade is inferred.

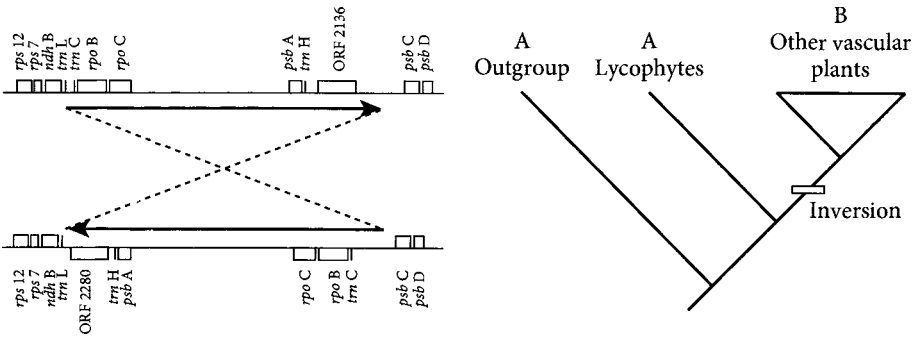


FIGURE A1.1 Example of phylogenetic inference from structural molecular data. A region of the plastid genome of land plants was found to exist in two distinct orientations. Orientation A was found in nonvascular land plants (the outgroup) and in lycophytes, whereas orientation B was found in all other vascular plants. This inversion implies that lycophytes are sister to all other vascular plants. Adapted from Raubeson and Jansen (1992).

An example of the latter approach is provided by a study of land plants (Figure A1.1). RFLP mapping approaches showed that all vascular plants except the clubmosses and their allies (the lycophytes) have a major inversion in their plastid genome. The arrangement in lycophytes is the same as that found in nonvascular land plants (mosses and liverworts). The presence of a structural molecular synapomorphy in land plants other than lycophytes shows that the lycophytes fall outside a clade that includes all other living vascular plants. Abundant subsequent research on land plant phylogeny has confirmed this conclusion. It seems that this inversion really did occur just once in the lineage leading to all living vascular plants except lycophytes.

Another classic example that fits this basic principle are chromosomal inversion phylogenies, which were reconstructed based on a logical analysis of a series of nested chromosomal inversions in Hawaiian fruit flies. Carson (1983) summarized more than a decade of such data and was able to provide a detailed phylogenetic tree for approximately 103 fly species—a tree that has largely been validated based on DNA sequence data (O'Grady et al. 2001).

Nonetheless, while structural molecular characters have provided definitive data in many cases, this does not mean that the Hennigian logic will always succeed. Our knowledge of molecular process is not good enough to definitively rule out independent origins of the same structural mutations. There-

fore, even when clades are supported by supposedly rare structural mutations, biologists still hope to corroborate those clades through the use of other kinds of data.

MORPHOLOGY AND OTHER MORPHOTYPIC DATA

Morphological features that reflect some underlying aspects of the genotype, making them heritable, can be used for inferring phylogenetic relationships. The basic approach with such data is to examine the variation among the taxa and develop an encoding scheme for summarizing that variation with a number of discrete states. As discussed in Appendix 2, the delimitation of characters and character states can be tricky, depending on a series of somewhat ambiguous judgment calls.

There are a number of kinds of data that are not morphological in the strict sense, but nonetheless contend with similar problems of character encoding. These are sometimes called "phenotypic data," but we do not find that term appropriate given that all observable features of organisms are technically phenotypes. Rather, we will use the term ***morphotypic*** to refer to those kinds of data for which character encoding is not defined *a priori*, but is guided by the observed variation among the taxa combined with insights into the homology and independence of the traits. Below is a list of some kinds of morphotypic data that have been used for phylogenetic analysis. It is common for a data matrix to include more than one kind of morphotypic data.

a. Morphology: Internal and external physical features of organisms.
b. Molecular morphology: The shape of particular molecules or molecular complexes, for example, ribosomes. Sometimes this variation can be captured effectively using sequence data, but occasionally the shape of subcellular structures has been treated as morphotypic data.
c. Development: How traits change during the lifetime of an organism.
d. Behavior: How organisms tend to behave. Mating behaviors in particular have been widely used.
e. Secondary biochemistry: Compounds that organisms accumulate and chemical reactions that its cells can conduct. The production of defensive secondary chemicals has proven important in plants, whereas fungi are often scored for the ability to cause a color reaction given a particular

substrate. In some cases the biochemical trait corresponds to the presence or absence of a particular gene, in which case this kind of data blends into structural molecular or molecular presence/absence data.

f. Biogeography: The geographic distribution of organisms can theoretically be scored like other morphotypic traits. More commonly, geographic history is mapped onto phylogenies that have been inferred from other data.

g. Ecology: The preferred habit or way of life of organisms, which is to say, aspects of their ecological niche. Examples include pollinator identity, prey type, or preferred habitat type.

FREQUENCY DATA

If different individuals scored from the same taxon vary in a trait, one strategy (the usual one nowadays) is to score that taxon as ***polymorphic***, i.e., containing multiple states, but not to worry about the frequency of each variant within the taxon. However, in earlier times, the frequency of a polymorphic character, usually an isozyme or RFLP allele, was considered to provide evidence of relatedness. A data matrix would be constructed in which the entries were not the presence or absence of a state, but the frequency of different alleles in different taxa. An example is shown in Figure A1.2. You will see that for each locus, the sum of the frequencies of each allele adds up to 1.0.

Methods have been developed for analyzing frequency data, most commonly involving first converting the frequency matrix into a distance matrix. However, when taxa share polymorphisms (e.g., each contains both alleles A and B of the same locus), then it is questionable whether the taxa are monophy-

Taxon list	Locus 1			Locus 2		Locus 3					
	1a	1b	1c	2a	2b	3a	3b	3c	3d	3e	Alleles
A	0.0	0.9	0.1	0.0	1.0	1.0	0.0	0.0	0.0	0.0	
B	0.0	0.3	0.7	0.2	0.8	0.4	0.6	0.0	0.0	0.0	
C	0.1	0.4	0.5	0.1	0.9	0.1	0.8	0.1	0.0	0.0	
D	0.9	0.1	0.0	0.8	0.2	0.0	0.1	0.0	0.9	0.0	
E	0.7	0.3	0.0	1.0	0.0	0.0	0.0	0.0	0.7	0.3	
F	1.0	0.0	0.0	0.9	0.1	0.0	0.1	0.1	0.4	0.4	

FIGURE A1.2 Example data set listing the frequency of alleles at three loci.

letic entities that can meaningfully be assigned to a single tip of the tree of life. This is one reason why it is now rare for frequency data to be used for phylogeny reconstruction.

DISTANCE DATA

The data types discussed so far are initially scored in the form of a character state matrix. Although they can be converted into a distance matrix (Chapter 8), they start out as a list of discrete states scored for each taxon. In contrast, two kinds of data that have been used for phylogenetic analysis, immunological cross-reactivity and DNA-DNA hybridization affinity, are collected in the form of a distance between a pair of taxa.

Immunological methods were popular in animals in the 1970s and 1980s. The intensity of the reaction between the immune serum of one animal with antigenic proteins from another animal was scored quantitatively in the laboratory. The underlying logic was that the greater the time since common ancestry the greater the protein differences, which should in turn lead to a more intense immune reaction.

Similarly, mainly in the 1980s, many systematists put great stock in DNA-DNA hybridization data as a measure of the overall sequence similarity of a pair of genomes. A typical experiment would involve single-stranded DNA from one taxon being attached to a column and then being allowed to hybridize with single-stranded (and radioactively labeled) DNA from another species. The greater the sequence similarity of the two genomes the more tightly would the complementary sequences bind. Overall genomic similarity could then be measured by looking at the release of radioactivity when the column was gradually heated up to melt apart the two strands.

Immunological and DNA-DNA distance measures can only be analyzed using distance methods, which is somewhat limiting. Also, whole genome distance data do not allow you to detect the existence of different gene trees for different parts of the genome (see Chapter 6). However, the main reason that these data are no longer used for phylogenetic research is that compared to DNA sequencing, the data are harder to generate and less readily repeatable. Nonetheless, in many (but not all) cases the conclusions reached using immunological or DNA-DNA hybridization data have since been validated through the analysis of DNA sequence data.

APPENDIX TWO

Generating a Morphological Data Matrix

Although the field of phylogenetics was founded on morphological data, most modern research uses molecular data, especially DNA sequences, which are abundant and relatively inexpensive to obtain. However, fossils can only be analyzed through the use of morphological data. Additionally, even when trees are inferred from molecular data, it often becomes necessary to build a morphological data matrix in order to study the evolution of morphological traits (Chapters 4 and 10). Therefore, despite the preeminence of molecular phylogenetics, it is important to know how morphological data are scored and assembled into data matrices.

Two steps can be recognized in the building of a morphological matrix, which we will call ***character encoding*** and ***character scoring.*** Character encoding involves deciding on the limits of characters and on the alternative states that are recognized for each character. For example, when you decide to score fur color using two states, brown/black and white, you have encoded one character (fur color). Character scoring involves looking at each taxon and assigning it a state for each encoded character. For example, you could score an otter as having brown/black fur. In practice observations made while scoring taxa often result in changes to character encoding, but it is still useful to distinguish these two steps.

Once a set of taxa has been selected for study, a systematist generally starts by looking for characteristics that appear to vary among them. Notice that, whereas many characters in a DNA sequence data matrix may be invariant, the way that morphological characters are selected means that constant characters are usually excluded from the outset.

Once some variation has been noted, the next challenge is to appropriately encode the characters. This is not straightforward. For example, imagine that

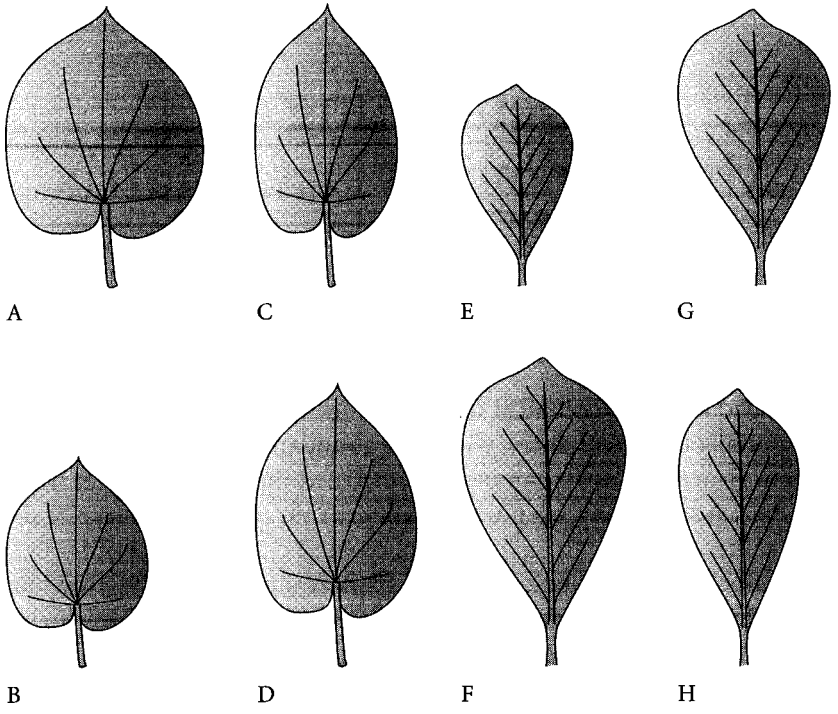


FIGURE A2.1 **A hypothetical example of variation in leaf shape.**

you observed that leaf shape and size differed among a set of eight plant species, as shown in Figure A2.1. How would you capture this variation?

Consider two of the numerous possible ways to encode this variation. (1) You recognize two basic leaf shapes, cordate (with the widest point in the lower half) and obcordate (with the widest point in the upper half), and two size classes. (2) You encode leaf length, leaf width, and the height of the widest point. The data matrix that might result from these two encoding schemes is shown in Tables A2.1 and A2.2. It is possible that the differences between these encoding schemes could alter the phylogenetic conclusions reached.

The decision among alternative encoding schemes is guided by a few, potentially conflicting, considerations. You want to capture as much of the variation as possible without "double counting." Scoring the same basic variation multiple times results in overweighting that variation to the point where it will

TABLE A2.1 One possible coding scheme for leaf shapes in Figure A2.1

Taxon	Leaf shape (0 = cordate; 1 = obcordate)	Leaf size (0 = small; 1 = large)
A	0	1
B	0	0
C	0	1
D	0	1
E	1	0
F	1	1
G	1	1
H	1	0

TABLE A2.2 An alternate possible coding scheme for leaf shapes in Figure A2.1.

Taxon	Leaf length (0 = short; 1 = long)	Leaf width (0 = wide; 1 = narrow)	Height of widest point (0 = below middle; 1 = above middle)
A	1	0	0
B	0	0	0
C	1	0	0
D	1	0	0
E	0	1	1
F	1	0	1
G	1	0	1
H	1	1	1

dominate the phylogenetic results. For example, you might be concerned that by measuring both length and width of these leaves you might score one basic trait, leaf size, twice.

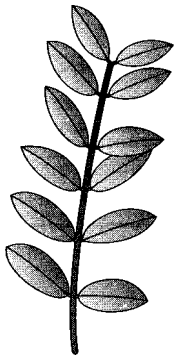


FIGURE A2.2 **A compound leaf.**

Another important consideration is that the character states recognized should really be versions of the same character. This is not always easy to decide. Suppose that close relatives of these plants have compound leaves (Figure A2.2). Should their "leaf shape" be encoded based on the individual leaflets or the outline of the whole compound leaf? The answer to this question will depend on your understanding of the developmental changes that resulted in a transition from compound to simple leaves, or vice versa.

Once characters are defined, the next question is how to delimit character states. If the variation is rather discrete between taxa, with little variation within taxa, as illustrated by leaf shape in the preceding example, then it may be easy to delimit character states. However, most morphological traits are inherently continuous and variation within taxa is common. Thus, it can be difficult to divide continuous variation into the discrete states needed for phylogenetic analysis. To illustrate the challenge, Figure A2.3 shows hypothetical data on leaf length in 10 species.

You might see the data in Figure A2.3 as being composed of three "clusters" of taxa corresponding to three states: small (A and E), medium (D, F, and H), and large (B, G, and I). In that case, you might score C as polymorphic for small-plus-medium and J as polymorphic for medium-plus-large. Alternatively, you could recognize two classes (small and large), or five size classes (A and E; C; D, F, and H; J; B, G, and I). Unfortunately, there is no well-grounded theory to tell you which of these encoding schemes will yield the best estimates of the phylogeny of these species.

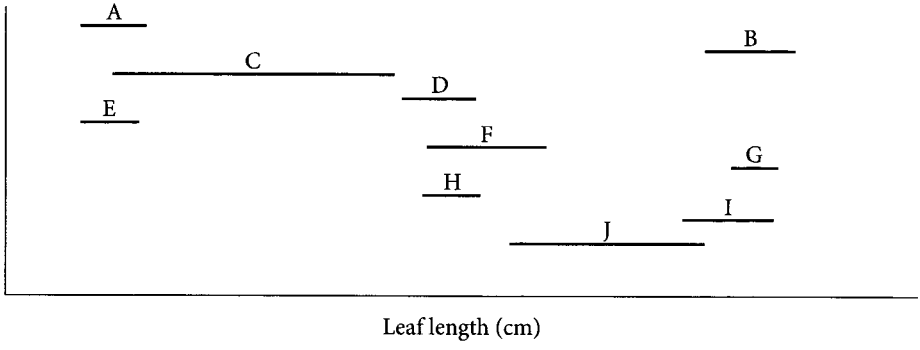


FIGURE A2.3 Variation in leaf length within and between 10 hypothetical taxa (A–J).

Taken together, you can probably see that there are many somewhat subjective decisions that must be made in encoding morphological data and that these are likely to be adjusted by observations made while scoring individual taxa. While subjectivity is something that makes scientists uncomfortable, this fact does not invalidate morphology as a source of phylogenetic data. So long as different encoding schemes capture the actual variation among taxa, then they should yield similar, if not identical, estimates of the tree. It is considered good practice to try a few different schemes and see if the phylogenetic conclusions remain the same. If they do, and if statistical approaches such as the nonparametric bootstrap (Chapter 9) indicate high support for clades, it is possible to achieve quite high confidence in the conclusions of a morphological phylogenetic analysis. For example, the relationships of many extinct groups that are known only from fossils are now known with a high degree of certainty.

Answers to Chapter Quizzes

CHAPTER 2 QUIZ

1. c
2. d
3. c [Cacti are well adapted to the desert environment but only occur in deserts in the New World. This pattern is only readily explicable if cacti radiated from an ancestor that lived in the New World.]
4. e
5. a [Humans are members of the ape clade. All other statements imply that some living species are more primitive or "lower" than others.]
6. d
7. c
8. e [Both evolutionary systematists and phylogenetic systematists agreed that evolutionary relatedness is an important consideration in taxonomy and both aspired to reconstruct phylogenetic trees. However, phylogeneticists considered relatedness to be the sole determinant of taxonomic status, whereas evolutionary systematists considered similarity to be important as well.]
9. e
10. a
11. A transitional fossil is a form that manifests some but not all of the derived features of a distinctive clade. When a group of organisms has many trait differences from all other known groups, one can predict that at one time there lived taxa that had some but not all of those derived traits. The discovery of such transitional forms corroborates this prediction. By so doing, transitional fossils support the evolutionary model by showing that traits of living organisms arose at different times in evolutionary history.

12. Tree thinking holds that different species living at the same time are connected to one another via common ancestors. Ladder thinking holds that some contemporaneous species are ancestral/primitive, whereas others are advanced/derived. Under tree thinking all contemporaneous taxa are equally evolved, whereas ladder thinking implies that some are more advanced than others. Additionally, ladder thinking suggests that there is one privileged direction in which evolution tends to proceed, implying that the direction of evolution is predetermined. In contrast, tree thinking implies that evolution is occurring simultaneously in many different "directions," and there are no intrinsic tendencies for progress along certain lines.

13. Lamarck concluded that organisms have an internal tendency or drive to become more advanced. Over generations, species would gradually improve themselves, moving upward on the *Scala Naturae*. The "gaps" at the bottom of the *Scala Naturae* would be filled by the spontaneous generation of new life. The mechanism for change was the inheritance of acquired characters (organisms strive to collect food, find mates, etc., causing their bodies to change, and these changes are passed on to offspring). Darwin held that there were very few origins or only one origin of life, and that the diversity of living species is explained by lineages descending from common ancestors. Thus, in Darwin's view, living species are not ancestors of other living species, but share common ancestors that lived in the past. While Darwin did not rule out the inheritance of acquired characters, he stressed natural selection as the primary mechanism by which evolution happens (heritable variation arises in an undirected manner and those new variants that improve organisms' ability to reproduce tend to increase in frequency).

14. Phylogenetic systematists would recognize that relatedness is determined by evolutionary kinship—recency of common ancestry—whereas similarity is determined by the extent to which visible traits have evolved in the time since common ancestry. There has been relatively little evolution in external form along the lineages leading to lizards and crocodiles, explaining why they are similar today. In contrast, there was a great deal of morphological evolution on the lineage leading to birds, causing birds to look very different from both lizards and crocodiles. Traits shared by crocodiles and lizards but not by birds were inherited by crocodiles and lizards in an unmodified form from their common ancestor but underwent evolutionary changes on the lineage leading to birds. However, crocodiles and birds share some features that they inherited from their common ancestor and these show that they are more closely related to each other than either is to lizards.

15. First, the use of the terms "higher" and "lower" connote a ladder of life, with species organized along the rungs, instead of a tree with species at the tips. This statement says that living species, like mosses and liverworts, gave rise to or are

ancestors of other living species, in this case, flowering plants. However, the correct statement is that these species have diverged from a common ancestor, which was not a moss, liverwort, or flowering plant. The statement also suggests that the "lower" organisms are somehow stuck, that they haven't evolved as much as other groups. However, both mosses and flowering plants have been evolving for the same period of time since their common ancestor and have both accumulated changes along their branches.

CHAPTER 3 QUIZ

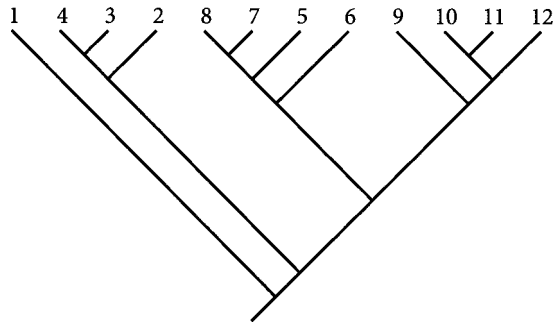
1. d 2. a 3. a 4. a 5. d 6. d

7. c 8. e 9. a 10. e 11. d

12. Upward (i.e., with ancestors at the bottom). This is clear because only in this direction does every snail have exactly two parents.

13. Answers vary.

14.



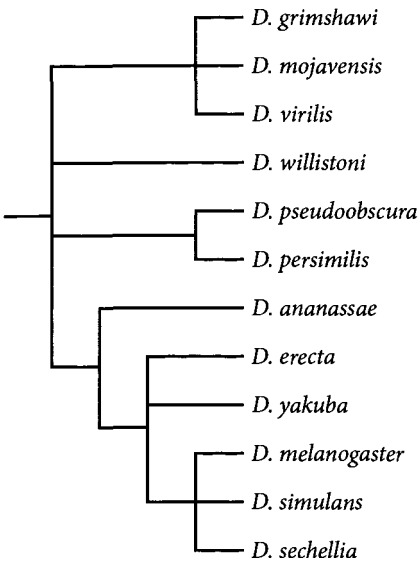
15. *D. grimshawi* [This tip is the furthest to the left showing that it has accumulated the least changes during its history.]

16. *D. ananassae* [This tip is furthest to the right.]

17. *D. sechellia, D. simulans,* and *D. melanogaster*

18. Of the clades annotated with bootstrap values, the least supported is (*D. virilis, D. mojavensis*) with 59%.

19.



20. One. [The RatCOX2 + MouseCOX2 subtree has moved from the external branch, HumanCOX2, to the internal branch separating HumanCOX1, MouseCOX1, and SheepCOX1 from the remaining taxa.]

CHAPTER 4 QUIZ

1. c 2. d 3. d 4. a 5. b 6. d 7. e

8. b [This is the only scenario in which the changes marked would agree with each other and would give rise to the indicated tip states.]

9. a 10. d 11. c 12. d

13. From left to right:

```
a a a c g a g g t t g g t c g c t t a a
g c a c g a g g c t c g t c g c t t a a
g g a c g g g a c t c g c g g c t t a g
a g a t c a g t t c c a t c a a t c c a
a g t c g a g t t t g a t c a a t t c a
```

14. If two traits are inferred by parsimony to have evolved on the same branch, then there is no way to infer which of them evolved first. Therefore, it is possible that the author of this figure selected randomly whether to mark the diet or beak character first. If, however, you knew of a species that was sister to bird B, which ate grubs but had a short beak, then the ordering of trait evolution shown would be the most parsimonious.

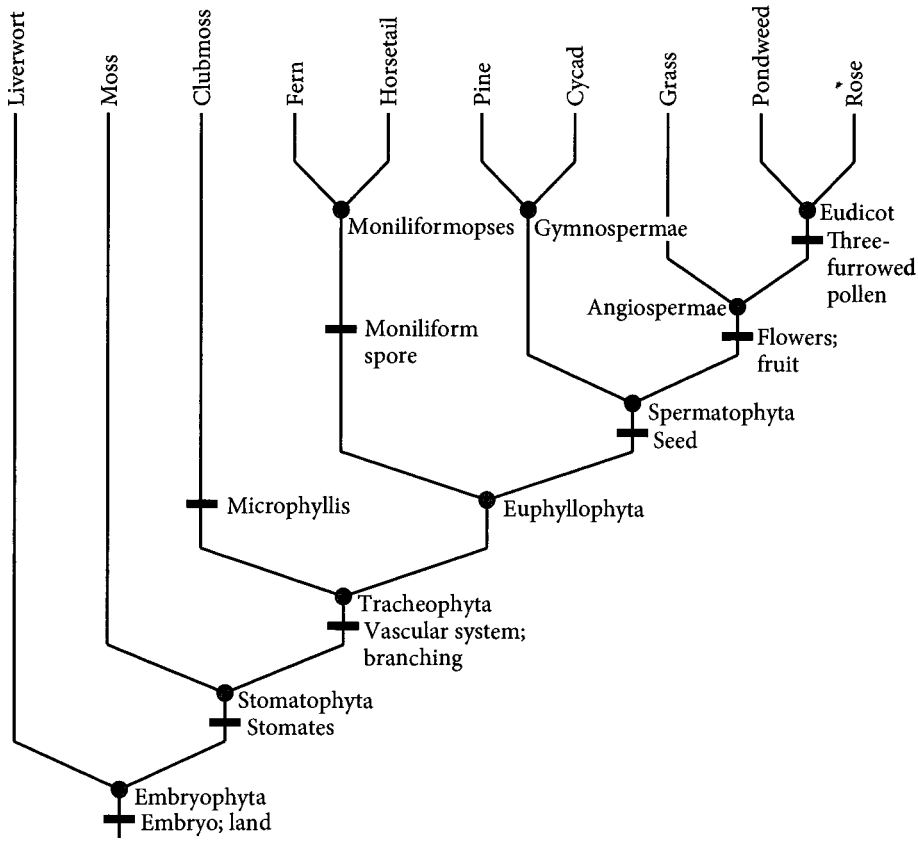
15. The line represents the transition from a population fixed for an ancestral character state to one fixed for a derived state. The character states are assumed to be encoded by alternative alleles at one or more genetic loci. At some point a mutation in one of these loci arose that contributed to the derived character state. Then, over many generations, the mutant (derived) allele increased in frequency until it became fixed, meaning that all individuals carried the derived allele and no individuals carried the ancestral allele. A particular character state change may involve changes at multiple loci. It is reasonable to represent the change with a single line so long as all the changes required to express the derived character went to fixation before subsequent lineage branching occurred.

16. A homologous trait is one inherited from a common ancestor. We could use the time machine to go back generation by generation along the human lineage, keeping track of which structure is homologous to the modern human thumb. When we get back to the last common ancestor of a human and a dog, we would note the structure of that ancestor that is homologous to the thumb. We would repeat the procedure while moving down the dog lineage. If the structure in the common ancestor that is homologous to the dewclaw is also homologous to the thumb, then the dewclaw and thumb are shown to be homologous.

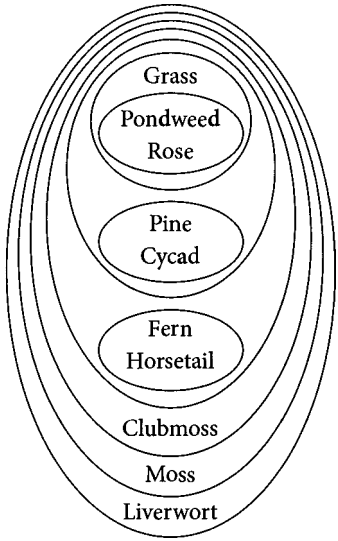
CHAPTER 5 QUIZ

1. a	2. c	3. c	4. a	5. c
6. e	7. b	8. e	9. a	10. b

11.


Liverwort
Moss
Clubmoss
Fern
Horsetail
Pine
Cycad
Grass
Pondweed
Rose
Moniliformopses
Gymnospermae
Eudicot
Three-furrowed pollen
Angiospermae
Moniliform spore
Flowers; fruit
Spermatophyta
Seed
Microphyllis
Euphyllophyta
Tracheophyta
Vascular system; branching
Stomatophyta
Stomates
Embryophyta
Embryo; land

12.



13. The traits that gorillas and chimpanzees share are primarily ones that were present in the last common ancestor of gorillas, humans, and chimpanzees (they are plesiomorphies). As such they do not contradict the molecular data, which show that humans and chimpanzees share a more recent common ancestor with each other than either shares with the gorilla. The differences between chimpanzees and humans arose because of rapid phenotypic evolution on the human lineage—at least for traits that are apparent to humans. This is why chimpanzees appear more similar to gorillas, despite being more closely related to humans.

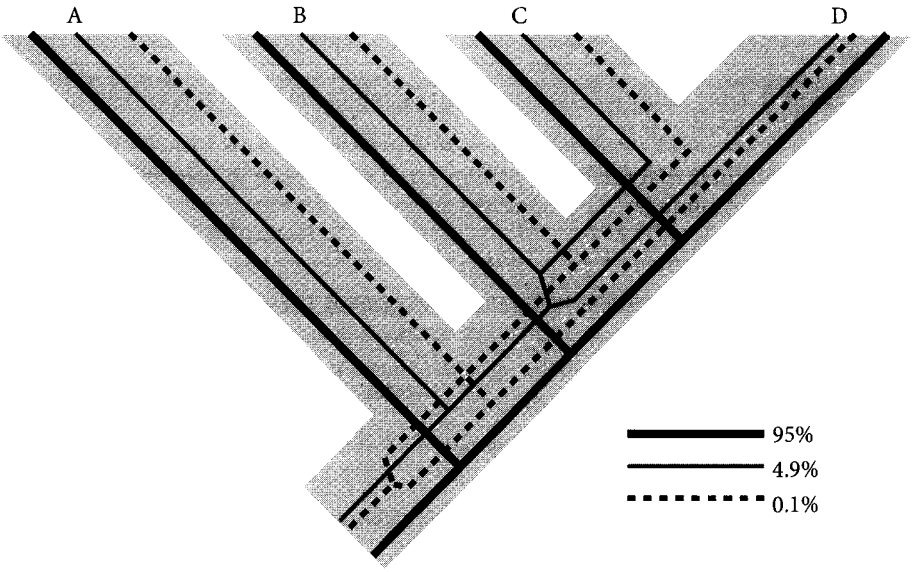
14. If some dinosaurs are more closely related to birds than to other dinosaurs, then dinosaurs would be a non-monophyletic (paraphyletic) group. Since classifications nowadays only include monophyletic groups, we need to include birds within the dinosaur clade. Consequently, if Dinosauria is to be a recognized taxon, a monophyletic group, birds must be included. In this case, dinosaurs are not extinct because some of their members (birds) continue to exist.

15. Traditional nomenclatural systems attach the names of taxa to whatever taxon at a specified rank contains the designated type. This means that even once one establishes that a taxon is monophyletic its name will be ambiguous without knowing its rank. This is problematic because rank is not an objective biological criterion. A phylogenetic nomenclatural system avoids this problem by attaching names directly to clades using multiple specifiers that are defined to be in or out of a group. This system ought to allow for greater stability for clade names, although it remains to be seen if it will replace the traditional rank-based codes of nomenclature.

CHAPTER 6 QUIZ

1. The contribution of a gamete from the ancestor below to create the diploid descendant above. Each individual receives one allele from each of its parents but can pass on its alleles to multiple individuals in the next generation.
2. a2
3. (((f1,b2),b1)c2)
4. ((a2,b2)(c2,g1))
5. h1
6. 2 (counting from the left, the 7th and 14th alleles)
7. d
8. c
9. a: 34%, b: 99%, c: 33%
10. c
11. b
12. c [one giving rise to the mammalian beta/gamma and delta/epsilon clade; one giving rise to the placental mammal beta and gamma copies]
13. a
14. c

15. a–b.



15. c. Because the two minor histories have uneven frequencies, it is likely that some phenomenon other than incomplete lineage sorting is responsible for the discordance. One possibility is that incomplete lineage sorting affected 0.1% of the genome, while about 4.8% of the genes experienced introgression between B and C. Another possibility is that 4.9% of the genes resulted from introgression between B and C, with one gene having experienced lateral gene transfer from D to B.

16. No, we should not expect all biological species to correspond to clades on the population tree. This is because the inability to interbreed with other populations is a derived trait. Given that not all populations will acquire reproductive isolation, some sets of populations will be united by the ability to interbreed but will not constitute a clade because other members of their clade have acquired reproductive isolation.

CHAPTER 7 QUIZ

1. d 2. c 3. a

4. c [Note that the outgroup has state 1 and therefore the fact that taxa A and E share state 0 implies that they are in a clade that excludes C.]

5. c [We can infer that the bases were inserted as opposed to deleted because they were not present in the outgroup (A).]

6. b 7. c 8. e 9. c 10. c

11. e [The first and third characters can be ignored because they are parsimony-uninformative. Of the remaining three characters, two have the same state in A + C and B + D. Therefore, the most parsimonious tree will be one that has either an (A,C) or a (B,D) clade, or both.]

12.

Taxa	1	2	3	4	5	6	7	8	9
A	T	G	T	G	A	A	C	A	A
B	T	G	T	G	A	C	C	A	A
C	T	G	C	G	G	C	C	T	A
D	A	G	C	G	G	C	G	T	A
E	A	A	C	T	A	A	G	T	G
F	A	A	C	T	A	A	G	C	G

Steps on 1:	1	1	1	1	2	2	1	2	1	Total Length = 12
Steps on 2:	2	1	2	1	1	1	2	3	1	Total Length = 14
Steps on 3:	1	1	1	1	2	2	1	2	1	Total Length = 12
Steps on 4:	3	2	2	2	2	1	3	3	2	Total Length = 20

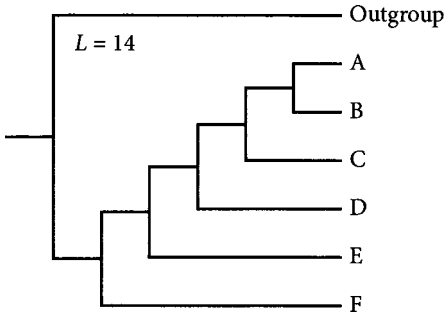
Note: Because tree 1 and tree 3 both correspond to the same unrooted tree topology (see Chapter 3), it is inevitable that they have the same tree length.

13. Sample answer.

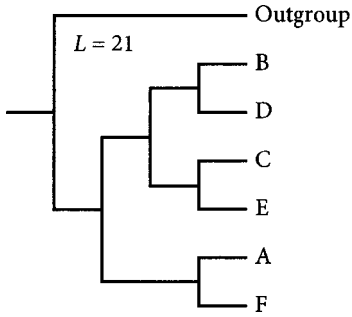
Taxon	1	2	3	4	5	6	7	8	9	10
A	A	C	A	-	-	A	A	T	C	T
B	A	C	A	G	G	A	A	G	A	G
C	A	C	T	G	G	A	G	G	A	G
D	A	T	T	G	G	A	G	G	A	T
E	G	T	T	G	G	A	T	G	A	G
F	G	T	T	C	G	A	T	A	C	G
Outgroup	G	T	T	C	G	G	T	A	C	T

- Characters 4 and 5 include gaps
- Characters 1–6 and 9–10 have one or two states; characters 7–8 have three or four states
- Characters 5 and 6 are uninformative
- Informative characters 1–4 and 7–8 are consistent with one another. Characters 9–10 conflict with some of the other six characters.

Tree 1



Tree 2



	Characters 1	2	3	4	5	6	7	8	9	10	
Steps on Tree 1:	1	1	1	1	0	1	2	2	2	3	Total Length = 14
Steps on Tree 2:	3	3	2	2	0	1	4	2	1	3	Total Length = 21

14. Imposing a weight of 20 on character 9 would result in a most parsimonious tree that has a (B,C,D,E) clade.

15. Taxon C is more closely related to D, despite having more character states in common with taxon B, that is, despite its greater similarity to B. This discrepancy is a result of many uniquely derived (autapomorphic) traits that are restricted to taxon D. This highlights the importance of focusing on shared derived characters and not overall similarity in inferring relationships.

16. In order for traits to provide evidence of evolutionary relationships, they need to be transmitted from ancestor to descendant species and they must exhibit some variation in the group being studied. When this applies, the observation that a set of taxa share the same derived character state provides evidence that those taxa form a clade, although that evidence may be contested by other traits. Characters that cannot group taxa together into clades (e.g., because they do not vary or because they vary only in one species) will not be informative for inferring phylogenies with parsimony.

17. The Hennigian method assumes that each derived trait evolves just once and never reverses to an ancestral state. Under these assumptions, some trees cannot explain the data found in the tips, meaning that they would be judged "false." Parsimony favors trees that invoke the fewest character state changes, which amounts to favoring the tree that maximizes the extent to which the Hennigian rules were followed. However, parsimony does not deny the possibility of homoplasy: even the longest tree is *possible* even though it is less consistent with the data.

18. With 100 taxa it is impossible to calculate the length of every tree. It will also almost certainly be impossible to use the branch-and-bound algorithm. Thus we are not assured of finding the most parsimonious tree. Nonetheless, using heuristic search algorithms, we have a good chance of finding the most parsimonious tree.

19. Equally weighted parsimony accords all characters in a matrix equal weight. This is not a lack of assumptions but an assumption that all characters are equally informative as to the true phylogeny. If we have evidence that some characters are likely to have evolved more rapidly than others and/or that we are more likely to have made errors in determining homology for some characters, then it is appropriate to accommodate this by character or character state weighting in generalized parsimony. Typically, we lower the weight of characters or character state transitions that are likely to show greater homoplasy. When our prior knowledge lacks any evidence that characters are more or less reliable, then equal weighting is justified.

20. False. Parsimony selects the tree that minimizes the amount of homoplasy (summed across characters), but that does not mean that the most parsimoni-

ous tree will have little or no homoplasy. So long as the characters have evolved at a slower rate than the rate of lineage branching, there are grounds for favoring trees with less homoplasy. This is true even in the case that all characters show some homoplasy. Parsimony assumes only that a reasonable number of characters evolve slowly, not that homoplasy is absent.

CHAPTER 8 QUIZ

1. c
2. d [Substituting 25,000,000 for t and 5.2×10^{-10} for μ into $\frac{1}{4}-\frac{1}{4}e^{-4/3\mu t}$ we get 0.0043. Thus, the site likelihood is 0.25 x 0.0043, or 0.0011.]
3. e [The probability that the first species has an A is 0.25 and the probability that it evolved into a G after 25 Ma (twice the time since common ancestry) is 0.0043.]
4. c 5. e 6. d 7. c 8. b 9. a 10. a, c, d
11. e [The probability of a singing frog being male, Pr(H|D), is equal to Pr(D|H), the probability of singing given that a frog is male (1.0) times Pr(H), the prior probability that a frog is male (0.4), divided by Pr(D), the probability of observing a singing frog. As indicated in the hint, the latter equals 0.46. Thus, the posterior probability that the frog is a male is 0.4/0.46 = 0.87.]
12. a, e
13. b [The ratio of the two likelihoods is $10^{-23}/10^{-22} = 0.1$.]
14. c [The posterior probability of a clade is the frequency with which that clade was sampled in the MCMC run. The seal + walrus clade appeared in 5 of 10 trees (Posterior probability (or clade credibility) = 5/10 = 0.5).]
15. The likelihood is the probability that a particular set of data would have arisen on a specific tree, while selecting all other parameters (e.g., branch lengths, ratio of transitions to transversions, etc.) such that the likelihood of the tree is maximized. To obtain this probability, we first have to lock in place a set of nuisance parameters and then calculate the probability that each individual character in the data set would have arisen in the observed pattern (summing over all possible ways that the character could have evolved). We then multiply these probabilities to obtain the probability of the entire data set arising on this tree (with these parameter values). We then search among possible parameter values to find the set that maximizes the likelihood. The probability of the data on the tree with these optimal parameter values is the likelihood score of the tree.

16. F81. [The four models listed are in order of complexity: JC assumes that bases are at the same frequency and all substitution types have the same rate; F81 is like JC but allows base frequencies to vary; HKY is like F81 but allows transitions and transversions to have different rates; GTR is like HKY but allows all six substitution types (two transitions and four transversions) to have different rates. Because parameters are added as you move down the list, the likelihood of each model must be equal to or higher than the model above it in the list. This means that the log-likelihood should get consistently higher (less negative) as you move down the list. F81 defies this pattern and, thus, must be the error.]

17. The neighbor-joining method can quickly calculate a tree (with branch lengths) for even a very large data set. This makes the method useful in situations where speed rather than accuracy matters. However, because the method just builds a single tree without providing a meaningful score of tree quality, the method does not allow one to evaluate competing trees. For example, it is not possible to identify the "second-best" tree with neighbor-joining. Additionally, neighbor-joining and other distance methods (including minimum evolution) lose accuracy by first converting a character state matrix, for example, aligned DNA sequences, into a distance matrix before using the latter as a basis for tree building.

18. The posterior probability of each tree is the likelihood multiplied by the prior probability of the tree, divided by the prior probability of the data. The latter quantity is the same for each tree and does not affect their relative posterior probabilities. If the prior probability of each tree is the same (that is, if we consider each tree to be equally probable *a priori*), then the relative probabilities are proportional to their relative likelihoods, that is, 2:1. We cannot calculate the absolute posterior probability of either tree, however, because we do not know the probability of these data (this would require knowing the prior probabilities and likelihoods of all trees).

19. Trees with higher posterior probability are visited more often because of the way that MCMC moves around tree space. At each step in the chain, the search has the possibility of moving to a new tree. MCMC will always walk "up" to a new tree that has a higher posterior probability. Although the MCMC can walk down to lower trees, its probability of doing so depends on the ratio of the posterior probability of the current and proposed trees. In this way, it will spend more time on trees with higher posterior probability and when we look over the whole MCMC walk, the parts of tree space that were visited the most will be the parts containing the trees with high posterior probability.

20. Unlike maximum likelihood, Bayesian analysis requires that the practitioner provide some prior information about the probability of the hypothesis (the tree and other parameters). Some consider this as biasing the results toward an expected outcome. Others view this as a useful property of Bayesian analyses, because in many cases we have information from other sources of data (geology, biology, etc.) that some hypotheses are more likely *a priori*. Another potential difference between the two approaches is the quantity being estimated. Those who favor Bayesian approaches might argue that we are not really interested in the probability of the data (given a tree) but in the probability of the tree itself, and Bayes' theorem (together with MCMC) provides a way to estimate that probability.

    Also, some might argue in favor of the Bayesian approach if they believe that it is a good idea to integrate over uncertainty in all of the parameters other than the tree topology, whereas others might take the contrary view and consider it preferable to find the optimal values of all nuisance parameters.

CHAPTER 9 QUIZ

1. d 2. a 3. d 4. d 5. e 6. d 7. e 8. a 9. c

10. Each character should contain the same number of each character state as the corresponding character in the original data matrix, but the assignment of taxa to character state could be different. Sample answer:

Taxon	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
A	A	T	A	C	C	A	A	A	C	T	G	C	A	T	A
B	A	G	G	T	T	A	T	A	A	T	G	C	G	G	G
C	A	T	G	C	C	T	G	G	C	T	G	T	A	C	A
D	A	T	A	T	C	T	T	G	C	T	A	T	A	G	A
E	A	G	G	C	C	A	A	A	A	T	A	C	G	C	A

11. Each character should correspond to one of the characters in the original data matrix, although characters from the original matrix can be represented more than once in the bootstrap pseudoreplicate. Sample answer:

Taxon	6	1	12	4	3	8	6	3	1	15	11	2	15	7	6
A	T	A	C	T	G	G	T	G	A	A	A	T	A	T	T
B	T	A	C	T	A	A	T	A	A	A	G	T	A	T	T
C	A	A	C	C	G	A	A	G	A	A	A	T	A	A	A
D	A	A	T	C	G	A	A	G	A	G	G	G	G	A	A
E	A	A	T	C	A	G	A	A	A	A	G	G	A	G	A

12. a. No. [Partition A probably contains fewer characters than the other partitions and all trees will tend to be shorter for this data set.]
    b. The sum of the most parsimonious trees for the three data sets is similar to the sum of lengths we would obtain by randomly assigning characters to the three partitions. This suggests that the characters in the three partitions have not tracked markedly different histories.

13. This is a parametric bootstrap because data were simulated along a tree using a model of sequence evolution and analysis of these simulated data sets was used to test a phylogenetic hypothesis. The result shows that even if the true tree was tree 2, we would be likely to infer that the true tree was tree 1. Therefore, the fact that most studies found tree 1 does not serve to refute tree 2.

14. Yes, *Rafflesia* is sister to Malpighiales (including *Viola*) with 100% or 1.0 posterior probability. Thus, trees with alternate relationships (e.g,. *Rafflesia* sister to *Oxalis*) were not visited during the MCMC chain and, assuming the chain was effectively exploring tree space, these other topologies have zero posterior probability. This conclusion would need to be modified, however, if the priors used in the Bayesian analysis were incorrect.

15. Deep coalescence events may result in a gene tree that conflicts with the species tree (see Chapter 6). Under certain scenarios, like rapid divergence of population lineages, gene tree/species tree discordance may be common due to incomplete lineage sorting. Nonetheless, if in the true species history A and B are more closely related than either is to C, we expect that more of the gene trees will show A and B as sisters than A and C or B and C as sisters. Thus, we usually (but not always) expect that the species tree that agrees with most of the gene trees is the correct

species tree. This will be the species tree that implies the fewest deep coalescence events.

16. We expect more genes to have a lion + leopard clade than a tiger + snow leopard clade. This is because the stem lineage of these two clades differ in inferred length: the tiger + snow leopard stem lineage is notably shorter in length. This should increase the chances of discordance due to incomplete lineage sorting. The confidence measures for these two clades might lead one to doubt this tree topology, but accepting that it *is* the tree topology, we expect more discordance about branches that are shorter.

CHAPTER 10 QUIZ

1. c 2. b 3. e 4. e 5. c 6. a 7. d 8. c 9. c

10. In scenario 2, the evolution of cone spines occurs in a lineage that was not being fed upon by birds. This contradicts the hypothesis that birds impose selective pressure for scale spines and suggests that the spines arose for some other reason. One species (Pine 3) in the group with spines later comes to be predated by birds. Even if the spines served to deter birds in this species, the spines would not be adaptive for bird defense because this was not their original function. Evolutionary biologists have termed such traits, whose current function is different than their original one, exaptations.

11. The realizations in the lower set of trees show greater support for correlated evolution because for each combination (00, 01, 10, 11) the difference between the observed and expected is greater. In a stochastic mapping framework, evidence for correlated evolution arises from the observation that two character states are present along the same branch more often than is expected, given the overall amount of branch length that each occupies across the tree.

12. If the trait is evolving by Brownian motion, then the path of future evolution is not affected by the history of trait evolution. Thus, the leaves could be above or below the current value of 6 cm with equal probability. One might be tempted to think that the data are incompatible with Brownian motion because there was a directional trend over the three time points, and Brownian motion is, by definition, not directional. However, the probability of an apparent trend over three time points by chance is high, 50%. Thus, the trend observed does not in itself provide a strong argument against Brownian motion.

13. Sample chronicle (should just list relationships and where particular traits evolved): The cetaceans, including *Basilosaurus*, are a clade of ungulates. They are most closely related to hippos with which they share an aquatic lifestyle. *Basilosaurus* shares two features with whales and dolphins: fins and carnivory. The loss of limbs occurred on the dolphin/whale lineage, following the split with *Basilosaurus*.

    Sample narrative (should provide a description of the selective forces that might have favored the evolution of the derived traits): The sister-group relationship between hippos and cetaceans suggests that aquatic living was the first step toward the invasion of the sea. The transition to feeding in water was followed by the evolution of carnivory and the evolution of fins, which may thus represent adaptations for swimming as opposed to walking. The hind limbs were later lost entirely, perhaps because an absence of hind limbs reduced drag and improved swimming ability.

14. In a phylogenetic context, we can detect significant correlated trait evolution when two character states appear together repeatedly in independent lineages. For example, when we look across the phylogeny of all flowering plants, we can see that wind pollination has arisen multiple times and this trait often co-occurs with particular flower traits (small size, no nectar). If there were no relationship between wind pollination and these flower traits, we would expect to see these traits arising just as often in lineages that are not wind pollinated. The "concentration" of shifts in flower traits in wind-pollinated lineages supports the hypothesis that these changes represent adaptations for wind pollination.

15. More closely related species tend, on average, to have more similar trait values than more distantly related species. In other words, the trait values for a given species *depend* at least to some extent on the values present in its ancestors. Standard statistical analyses like regression and correlation assume that the data points are *independent*. Comparative data (across species) violate this assumption. However, if we know the phylogeny and can assume a model of how traits evolve along it, we can use methods like independent contrasts to formally account for the shared history.

16. The concentrated changes test would only be preferred in scenarios in which branch lengths were unavailable for the group of organisms being studied. This might be the case if no phylogeny is available or if the phylogeny is "pieced together" from multiple studies that used different and noncombinable sources of data. However, when branch lengths are available, model-based methods have the advantage of incorporating the realistic assumption that changes in character states are more likely along longer branches, that is, over longer stretches of time. By contrast, the parsimony-based concentrated changes test assumes that change

is equally likely on all branches of the tree, regardless of their length. This would only be a safe assumption for very slowly evolving traits.

CHAPTER 11 QUIZ

1. c 2. a 3. c

4. d [The species occurring in each community tend to each be drawn from one of the three major clades. However, within those clades, co-occurring species are randomly distributed across the clade—sometimes including co-occurring sister groups.]

5. c [The species occurring in a community are drawn from all three clades. It appears that in many cases three very closely related species are distributed with one in each community.]

6. a [The species occurring in each community tend to each be drawn from one of the three major clades. However, within those clades, co-occurring species seem to be overdispersed, with few cases of co-occurring sister groups.]

7. c 8. a 9. c 10. a 11. d

12. We could test whether asexuality is a dead end by comparing the size of asexual clades to the most closely related sexual lineages. If asexuality is a dead end, we would predict that asexual clades would be less species-rich than their sexual sister groups, and the magnitude of the difference would help us decide the chances that such a size difference could have arisen by chance. It might be useful to examine multiple such sister-group pairs to determine the generality of this conclusion. However, any such sister-group comparison assumes that the character history is known; thus, a more robust test might apply the BiSSE model, which simultaneously estimates character transitions and the effect of those transitions on the rate at which lineages diversify.

13. Because the rate of going to state 1 (wind) is higher than the reverse, we expect transition rates to increase the number of wind-pollinated lineages over time. The diversification rate of wind-pollinated lineages ($\lambda_1 - \mu_1 = 0.2 - 0$) is twice that of animal-pollinated lineages. This too favors an increase in the number of wind-pollinated lineages. Thus, over time, we expect the numbers of animal-pollinated lineages to eventually go to zero. However, there is likely to be variation around this result, depending on the ancestral state at the start of the simulation and the length of time the simulation is allowed to run.

References

(page locations of references given in bold font)

Abouheif, E. 1997. Developmental genetics and homology: A hierarchical approach. *Trends in Ecology & Evolution* 12 (10):405–408. **100**

Ackerly, D. D. 2004. Adaptation, niche conservatism, and convergence: Comparative studies of leaf evolution in the California chaparral. *American Naturalist* 163 (5):654–671. **381**

Agnarsson, I., M. Kuntner, and L. J. May-Collado. 2010. Dogs, cats, and kin: A molecular species-level phylogeny of Carnivora. *Molecular Phylogenetics and Evolution* 54 (3):726–745. **202**

Alfaro, M. E., F. Santini, C. Brock, H. Alamillo, A. Dornburg, D. L. Rabosky, G. Carnevale, and L. J. Harmon. 2009. Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. *Proceedings of the National Academy of Sciences of the United States of America* 106 (32):13410–13414. **374–375**

Ané, C. 2010. Reconstructing concordance trees and testing the coalescent model from genome-wide data sets. In *Estimating species trees: Practical and theoretical aspects*, edited by L. L. Knowles and L. S. Kubatko. Hoboken, New Jersey: Wiley-Blackwell. **151, 295–296**

Ané, C., B. Larget, D. A. Baum, S. D. Smith, and A. Rokas. 2007. Bayesian estimation of concordance among gene trees. *Molecular Biology and Evolution* 24 (2):412–426. **297**

Archie, J. W. 1989. A randomization test for phylogenetic information from systematic data. *Systematic Zoology* 38 (3):239–252. **21–22, 30, 297**

Barkman, T. J., G. Chenery, J. R. McNeal, J. Lyons-Weiler, W. J. Ellisens, G. Moore, A. D. Wolfe, and C. W. dePamphilis. 2000. Independent and combined analyses of sequences from all three genomic compartments converge on the root of flowering plant phylogeny. *Proceedings of the National Academy of Sciences of the United States of America* 97 (24):13166–13171. **303**

Barkman, T. J., S. H. Lim, K. M. Salleh, and J. Nais. 2004. Mitochondrial DNA sequences reveal the photosynthetic relatives of *Rafflesia*, the world's largest flower. *Proceedings of the National Academy of Sciences of the United States of America* 101 (3):787–792. **303**

Baum, D. A. 2007. Concordance trees, concordance factors, and the exploration of reticulate genealogy. *Taxon* 56 (2):417–426. **166**

Baum, D. A. 2009. Species as ranked taxa. *Systematic Biology* 58 (1):74–86. **166**

Baum, D. A., and A. Larson. 1991. Adaptation reviewed—a phylogenetic methodology for studying character macroevolution. *Systematic Zoology* 40 (1):1–18. **340**

Baum, D. A., and S. Offner. 2008. Phylogenies and tree-thinking. *American Biology Teacher* 70 (4):222–229. **68, 133**

Baum, D. A., and K. L. Shaw. 1995. Genealogical perspectives on the species problem. *Monographs in Systematic Botany from the Missouri Botanical Garden* 53:289–303. **166**

Baum, D. A., S. D. Smith, and S. S. Donovan. 2005. The tree-thinking challenge. *Science* 310 (5750):979–980. **68**

Baum, D. A., S. D. Smith, A. Yen, W. S. Alverson, R. Nyffeler, B. A. Whitlock, and R. L. Oldham. 2004. Phylogenetic relationships of Malvatheca (Bombacoideae and Malvoideae; Malvaceae *sensu lato*) as inferred from plastid DNA sequences. *American Journal of Botany* 91 (11):1863–1871. **57**

Bessey, C. E. 1915. The phylogenetic taxonomy of flowering plants. *Annals of the Missouri Botanical Garden* 2 (1/2):109–164. **23–24**

Bossuyt, F., M. Meegaskumbura, N. Beenaerts, D. J. Gower, R. Pethiyagoda, K. Roelants, A. Mannaert, M. Wilkinson, M. M. Bahir, K. Manamendra-Arachchi, P. K. L. Ng, C. J. Schneider, O. V. Oommen, and M. C. Milinkovitch. 2004. Local endemism within the western Ghats-Sri Lanka biodiversity hotspot. *Science* 306 (5695):479–481. **382**

Boussau, B., and V. Daubin. 2010. Genomes as documents of evolutionary history. *Trends in Ecology & Evolution* 25 (4):224–232. **166**

Bowler, P. J. 2003. *Evolution: The history of an idea.* 3rd ed. Berkeley: University of California. **30**

Bremer, K. 1994. Branch support and tree stability. *Cladistics* 10 (3):295–304. **272**

Brooks, D. R. 1981. Hennig's parasitological method—a proposed solution. *Systematic Zoology* 30 (3):229–249. **381**

Brooks, D. R., D. A. McLennan, J. P. Carney, M. D. Dennison, and C. A. Goldman. 1994. Phylogenetic systematics: Developing an hypothesis of amniote relationships. In *Tested studies for laboratory teaching, Volume 15*, edited by C. A. Goldman. Toronto: Association for Biology Laboratory Education. **207**

Browne, J. 2006. *Darwin's Origin of the Species: A biography*, Crows Nest NSW, Australia: Allen & Unwin. **30**

Brown, J. M., and A. R. Lemmon. 2007. The importance of data partitioning and the utility of Bayes factors in Bayesian phylogenetics. *Systematic Biology* 56 (4):643–655. **297**

Butler, M. A., and A. A. King. 2004. Phylogenetic comparative analysis: A modeling approach for adaptive evolution. *American Naturalist* 164 (6):683–695. **340**

Carson, H. L. 1983. Chromosomal sequences and interisland colonizations in Hawaiian *Drosophila*. *Genetics* 103 (3):465–482. **394**

Carstens, B. C., and T. A. Dewey. 2010. Species delimitation using a combined coalescent and information-theoretic approach: An example from North American *Myotis* bats. *Systematic Biology* 59 (4):400–414. **294**

Carstens, B. C., and L. L. Knowles. 2007. Estimating species phylogeny from gene-tree probabilities despite incomplete lineage sorting: An example from *Melanoplus* grasshoppers. *Systematic Biology* 56 (3):400–411. **293**

Catley, K. M., and L. R. Novick. 2008. Seeing the wood for the trees: An analysis of evolutionary diagrams in biology textbooks. *Bioscience* 58 (10):976–987. **68**

Catley, K. M., and L. R. Novick. 2009. Digging deep: Exploring college students' knowledge of macroevolutionary time. *Journal of Research in Science Teaching* 46 (3):311–332. **68**

Cavender-Bares, J., D. D. Ackerly, D. A. Baum, and F. A. Bazzaz. 2004. Phylogenetic overdispersion in Floridian oak communities. *American Naturalist* 163 (6):823–843. **359, 381**

Cavender-Bares, J., K. H. Kozak, P. V. A. Fine, and S. W. Kembel. 2009. The merging of community ecology and phylogenetic biology. *Ecology Letters* 12 (7):693–715. **381**

Choi, S. C., and J. Hey. 2011. Joint inference of population assignment and demographic history. *Genetics* 189(2):561–77. **296**

Clusella-Trullas, S., J. S. Terblanche, T. M. Blackburn, and S. L. Chown. 2008. Testing the thermal melanism hypothesis: A macrophysiological approach. *Functional Ecology* 22 (2):232–238. **339**

Cooper, A., C. Lalueza-Fox, S. Anderson, A. Rambaut, J. Austin, and R. Ward. 2001. Complete mitochondrial genome sequences of two extinct moas clarify ratite evolution. *Nature* 409 (6821):704–707. **354–355**

Coyne, J. A. 2009. *Why evolution is true*. New York: Viking. **30**

Crisp, M. D., and L. G. Cook. 2005. Do early branching lineages signify ancestral traits? *Trends in Ecology & Evolution* 20 (3):122–128. **30**

Darwin, C. 1859. *On the origin of species by means of natural selection*. London: John Murray. **12, 14–15, 18–19, 21, 23, 120**

Darwin, E. 1794. *Zoonomia*. London: J. Johnson. **11–12**

Davis, B. W., G. Li, and W. J. Murphy. 2010. Supermatrix and species tree methods resolve phylogenetic relationships within the big cats, *Panthera* (Carnivora: Felidae). *Molecular Phylogenetics and Evolution* 56 (1):64–76. **304**

Dawkins, R. 2009. *The greatest show on earth: The evidence for evolution*. New York: Free Press. **30**

de Queiroz, K. 2005. Different species problems and their resolution. *Bioessays* 27 (12):1263–1269. **166**

de Queiroz, K., and J. Gauthier. 1990. Phylogeny as a central principle in taxonomy: Phylogenetic definitions of taxon names. *Systematic Zoology* 39 (4):307–322. **133**

de Queiroz, K., and J. Gauthier. 1992. Phylogenetic taxonomy. *Annual Review of Ecology and Systematics* 23:449–480. **133**

de Queiroz, K., and J. Gauthier. 1994. Toward a phylogenetic system of biological nomenclature. *Trends in Ecology & Evolution* 9 (1):27–31. **133**

de Wit, M., M. Jeffrey, H. Bergh, and L. Nicolaysen. 1999. Gondwana reconstruction and dispersion. *Search and Discovery* 30001. http://www.searchanddiscovery.com/documents/97019/index.htm **354**

Des Marais, D. L., and M. D. Rausher. 2008. Escape from adaptive conflict after duplication in an anthocyanin pathway gene. *Nature* 454 (7205): 762–765. **170**

Donoghue, M. J. 1985. A critique of the biological species concept and recommendations for a phylogenetic alternative. *Bryologist* 88 (3):172–181. **166**

Donoghue, M. J., R. G. Olmstead, J. F. Smith, and J. D. Palmer. 1992. Phylogenetic relationships of Dipsacales based on *rbc*L sequences. *Annals of the Missouri Botanical Garden* 79 (2):333–345. **297**

Doolittle, W. F. 1999. Phylogenetic classification and the universal tree. *Science* 284 (5423):2124–2128. **166**

Drosophila 12 Genomes Consortium. 2007. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 450 (7167):203–218. **74**

Drummond, A. J., S. Y. W. Ho, M. J. Phillips, and A. Rambaut. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biology* 4 (5):699–710. **368, 381**

Drummond, A. J., and A. Rambaut. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* 7:214. **366**

Dubois, A. 2007. Phylogeny, taxonomy and nomenclature: The problem of taxonomic categories and of nomenclatural ranks. *Zootaxa* (1519):27–68. **133**

Edwards, S. V., L. Liu, and D. K. Pearl. 2007. High-resolution species trees without concatenation. *Proceedings of the National Academy of Sciences of the United States of America* 104 (14):5936–5941. **297**

Ereshefsky, M. 1989. Where's the species? Comments on phylogenetic species concepts. *Biology & Philosophy* 4 (1):89–96. **166**

Ereshefsky, M. 2011. Mystery of mysteries: Darwin and the species problem. *Cladistics* 27 (1):67–79. **166**

Faith, D. P., and P. S. Cranston. 1991. Could a cladogram this short have arisen by chance: On permutation tests for cladistic structure. *Cladistics* 7 (1):1–28. **297**

Farris, J. S. 1974. Formal definitions of paraphyly and polyphyly. *Systematic Zoology* 23 (4):548–554. **133**

Farris, J. S. 1983. The logical basis of phylogenetic analysis. In *Advances in cladistics*, edited by N. Platnick and V. Funk. New York: Columbia Univ. Press. **207**

Farris, J. S. 1989. The retention index and the rescaled consistency index. *Cladistics* 5 (4):417–419. **100**

Farris, J. S. 2000. Corroboration versus "strongest evidence." *Cladistics* 16 (4):385–393. **207**

Farris, J. S., M. Källersjö, A. G. Kluge, and C. Bult. 1994. Testing the significance of incongruence. *Cladistics* 10 (3):315–319. **288, 297**

Felsenstein, J. 1973. Maximum likelihood and minimum-steps methods for estimating evolutionary trees from data on discrete characters. *Systematic Zoology* 22 (3):240–249. **244**

Felsenstein, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Systematic Zoology* 27 (4):401–410. **205, 207**

Felsenstein, J. 1981a. Evolutionary trees from DNA sequences: A maximum likelihood approach. *Journal of Molecular Evolution* 17 (6):368–376. **225, 259**

Felsenstein, J. 1981b. A likelihood approach to character weighting and what it tells us about parsimony and compatibility. *Biological Journal of the Linnean Society* 16 (3):183–196. **207**

Felsenstein, J. 1983. Parsimony in systematics—biological and statistical issues. *Annual Review of Ecology and Systematics* 14:313–333. **207**

Felsenstein, J. 1984. Distance methods for inferring phylogenies: A justification. *Evolution* 38 (1):16–24. **259**

Felsenstein, J. 1985a. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39 (4):783–791. **273, 297**

Felsenstein, J. 1985b. Phylogenies and the comparative method. *American Naturalist* 125 (1):1–15. **311, 340**

Felsenstein, J. 2004. *Inferring phylogenies*. Sunderland, Massachusetts: Sinauer. **1, 259**

Fitch, W. M. 1970. Distinguishing homologous from analogous proteins. *Systematic Zoology* 19 (2):99–113. **166**

Fitch, W. M., 1971. Toward defining the course of evolution: Minimum change for a specified tree topology. *Systematic Zoology* 20 (4):406-416. **202**

Fitch, W. M., and E. Margoliash. 1967. Construction of phylogenetic trees. *Science* 155 (760):279–284. **259**

FitzJohn, R. G., W. P. Maddison, and S. P. Otto. 2009. Estimating trait-dependent speciation and extinction rates from incompletely resolved phylogenies. *Systematic Biology* 58 (6):595–611. **381**

Flynn, J. J., and M. A. Nedbal. 1998. Phylogeny of the Carnivora (Mammalia): Congruence vs incompatibility among multiple data sets. *Molecular Phylogenetics and Evolution* 9 (3):414–426. **200**

Fraune, S., and M. Zimmer. 2008. Host-specificity of environmentally transmitted *Mycoplasma*-like isopod symbionts. *Environmental Microbiology* 10 (10):2497–2504. **386**

Friedman, J., and S. C. H. Barrett. 2008. A phylogenetic analysis of the evolution of wind pollination in the angiosperms. *International Journal of Plant Sciences* 169 (1):49–58. **323**

Garland, T., P. H. Harvey, and A. R. Ives. 1992. Procedures for the analysis of comparative data using phylogenetically independent contrasts. *Systematic Biology* 41 (1):18–32. **338–340**

Givnish, T. J., K. C. Millam, A. R. Mast, T. B. Paterson, T. J. Theim, A. L. Hipp, J. M. Henss, J. F. Smith, K. R. Wood, and K. J. Sytsma. 2009. Origin, adaptive radiation and diversification of the Hawaiian lobeliads (Asterales: Campanulaceae). *Proceedings of the Royal Society B-Biological Sciences* 276 (1656):407–416. **350**

Goldberg, E. E., J. R. Kohn, R. Lande, K. A. Robertson, S. A. Smith, and B. Igić. 2010. Species selection maintains self-incompatibility. *Science* 330 (6003):493–495. **380–381**

Goldman, N. 1993. Statistical tests of models of DNA substitution. *Journal of Molecular Evolution* 36 (2):182–198. **259**

Goldman, N., J. P. Anderson, and A. G. Rodrigo. 2000. Likelihood-based tests of topologies in phylogenetics. *Systematic Biology* 49 (4):652–670. **297**

Gould, S. J., and E. S. Vrba. 1982. Exaptation—a missing term in the science of form. *Paleobiology* 8 (1):4–15. **340**

Graur, D., and W. H. Li. 2000. *Fundamentals of molecular evolution*. 2nd ed. Sunderland, Massachusetts: Sinauer. **100, 166**

Gregory, T. R. 2008. Understanding evolutionary trees. *Evolution: Education and Outreach* 1:121–137. **68**

Hackett, S. J., R. T. Kimball, S. Reddy, R. C. K. Bowie, E. L. Braun, M. J. Braun, J. L. Chojnowski, W. A. Cox, K. L. Han, J. Harshman, C. J. Huddleston, B. D. Marks, K. J. Miglia, W. S. Moore, F. H. Sheldon, D. W. Steadman, C. C. Witt, and T. Yuri. 2008. A phylogenomic study of birds reveals their evolutionary history. *Science* 320 (5884):1763–1768. **354**

Hall, B. G. 2011. *Phylogenetic trees made easy*. 4th ed. Sunderland, Massachusetts: Sinauer. **255**

Hall, B. K. 2003. Descent with modification: The unity underlying homology and homoplasy as seen through an analysis of development and evolution. *Biological Reviews* 78 (3):409–433. **100**

Harvey, P. H., and M. D. Pagel. 1991. *The comparative method in evolutionary biology*. Oxford: Oxford University Press. **340**

Hasegawa, M., H. Kishino, and T. A. Yano. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution* 22 (2):160–174. **228**

Heled, J., and A. J. Drummond. 2010. Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution* 27 (3):570–80. **297**

Helmus, M. R., K. Savage, M. W. Diebel, J. T. Maxted, and A. R. Ives. 2007. Separating the determinants of phylogenetic community structure. *Ecology Letters* 10 (10):917–925. **361**

Hennig, W. 1966. *Phylogenetic systematics*. Urbana, Illinois: University of Illinois Press. **27, 30, 100, 133, 178, 207**

Hibbett, D. S. 2001. Shiitake mushrooms and molecular clocks: Historical biogeography of *Lentinula*. *Journal of Biogeography* 28 (2):231–241. **364–365**

Hickman, C. P. Jr., L. S. Roberts, S. L. Keen, D. J. Eisenhour, A. Larson, and H. I'Anson. 2011. *Integrated principles of zoology*. 15th ed. Boston: McGraw-Hill Higher Education. **123**

Hillis, D. M., and J. J. Bull. 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Systematic Biology* 42 (2):182–192. **297**

Hillis, D. M., J. J. Bull, M. E. White, M. R. Badgett, and I. J. Molineux. 1992. Experimental phylogenetics: Generation of a known phylogeny. *Science* 255 (5044):589–592. **205**

Hillis, D. M., and J. P. Huelsenbeck. 1992. Signal, noise, and reliability in molecular phylogenetic analyses. *Journal of Heredity* 83 (3):189–195. **297**

Hillis, D. M., and J. P. Huelsenbeck. 1994. Support for dental HIV transmission. *Nature* 369 (6475):24–25. **6**

Hoberg, E. P., N. L. Alkire, A. de Queiroz, and A. Jones. 2001. Out of Africa: Origins of the *Taenia* tapeworms in humans. *Proceedings of the Royal Society B-Biological Sciences* 268 (1469):781–787. **356–357**

Hodges, S. A. 1997. Floral nectar spurs and diversification. *International Journal of Plant Sciences* 158 (6):S81–S88. **377–378**

Hodges, S. A., and M. L. Arnold. 1995. Spurring plant diversification: Are floral nectar spurs a key innovation? *Proceedings of the Royal Society B-Biological Sciences* 262 (1365):343–348. **378**

Hofmann, C. M., T. W. Cronin, and K. E. Omland. 2006. Using spectral data to reconstruct evolutionary changes in coloration: Carotenoid color evolution in New World orioles. *Evolution* 60 (8):1680–1691. **335**

Holder, M., and P. O. Lewis. 2003. Phylogeny estimation: Traditional and Bayesian approaches. *Nature Reviews Genetics* 4 (4):275–284. **259**

Howarth, D. G., and D. A. Baum. 2005. Genealogical evidence of homoploid hybrid speciation in an adaptive radiation of *Scaevola* (Goodeniaceae) in the Hawaiian Islands. *Evolution* 59 (5): 948–961. **167**

Huelsenbeck, J. P. 1997. Is the Felsenstein zone a fly trap? *Systematic Biology* 46 (1):69–74. **282–283, 297**

Huelsenbeck, J. P., and J. J. Bull. 1996. A likelihood ratio test to detect conflicting phylogenetic signal. *Systematic Biology* 45 (1):92–98. **297**

Huelsenbeck, J. P., and K. A. Crandall. 1997. Phylogeny estimation and hypothesis testing using maximum likelihood. *Annual Review of Ecology and Systematics* 28:437–466. **259**

Huelsenbeck, J. P., B. Larget, and M. E. Alfaro. 2004. Bayesian phylogenetic model selection using reversible jump Markov chain Monte Carlo. *Molecular Biology and Evolution* 21 (6):1123–1133. **259**

Huelsenbeck, J. P., B. Larget, R. E. Miller, and F. Ronquist. 2002. Potential applications and pitfalls of Bayesian inference of phylogeny. *Systematic Biology* 51 (5):673–688. **259**

Huelsenbeck, J. P., R. Nielsen, and J. P. Bollback. 2003. Stochastic mapping of morphological characters. *Systematic Biology* 52 (2):131–158. **340, 346**

Huelsenbeck, J. P., and B. Rannala. 1997. Phylogenetic methods come of age: Testing hypotheses in an evolutionary context. *Science* 276 (5310):227–232. **297**

Huelsenbeck, J. P., B. Rannala, and Z. H. Yang. 1997. Statistical tests of host-parasite cospeciation. *Evolution* 51 (2):410–419. **381**

Hull, D. L. 1970. Contemporary systematic philosophies. *Annual Review of Ecology and Systematics* 1:19–54. **133**

Hull, D. L. 1988. *Science as a process: An evolutionary account of the social and conceptual development of science.* Chicago: University of Chicago Press. **30**

Humphries, C. J., and L. R. Parenti. 1999. *Cladistic biogeography: Interpreting patterns of plant and animal distributions.* 2nd ed. Oxford: Oxford University Press. **381**

Johnson, W. E., E. Eizirik, J. Pecon-Slattery, W. J. Murphy, A. Antunes, E. Teeling, and S. J. O'Brien. 2006. The late Miocene radiation of modern Felidae: A genetic assessment. *Science* 311 (5757):73–77. **351**

Jukes, T. H., and C. R. Cantor. 1969. Evolution of protein molecules. In *Mammalian protein metabolism*, edited by H. N. Munro. New York: Academic Press. **224**

Kishino, H., and M. Hasegawa. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. *Journal of Molecular Evolution* 29 (2):170–179. **297**

Kitching, I. J., P. L. Forey, C. J. Humphries, and D. M. Williams. 1998. Measures of character fit and character weighting. In *Cladistics: The theory and practice of parsimony analysis*, edited by I. J. Kitching, P. L. Forey, C. J. Humphries, and D. M. Williams. Oxford: Oxford University Press. **100**

Knowles, L. L., and B. C. Carstens. 2007. Estimating a geographically explicit model of population divergence. *Evolution* 61 (3):477–493. **297**

Krause, J., T. Unger, A. Nocon, A. S. Malaspinas, S. O. Kolokotronis, M. Stiller, L. Soibelzon, H. Spriggs, P. H. Dear, A. W. Briggs, S. C. E. Bray, S. J. O'Brien, G. Rabeder, P. Matheus, A. Cooper, M. Slatkin, S. Pääbo, and M. Hofreiter. 2008. Mitochondrial genomes reveal an explosive radiation of extinct and extant bears near the Miocene-Pliocene boundary. *BMC Evolutionary Biology* 8:220. **386**

Kubatko, L. S., B. C. Carstens, and L. L. Knowles. 2009. STEM: Species tree estimation using maximum likelihood for gene trees under coalescence. *Bioinformatics* 25 (7):971–973. **294, 297**

Lamarck, J. B. 1809. *Philosophie Zoologique.* Translated by H. Elliot. New York: Macmillan and Company. **10–11**

Larget, B., and D. L. Simon. 1999. Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees. *Molecular Biology and Evolution* 16 (6):750–759. **259**

Lau, S. K. P., P. C. Y. Woo, K. S. M. Li, Y. Huang, H. W. Tsoi, B. H. L. Wong, S. S. Y. Wong, S. Y. Leung, K. H. Chan, and K. Y. Yuen. 2005. Severe acute respiratory syn-

drome coronavirus-like virus in Chinese horseshoe bats. *Proceedings of the National Academy of Sciences of the United States of America* 102 (39):14040–14045. **6**

Laurin, M. 2008. The splendid isolation of biological nomenclature. *Zoologica Scripta* 37 (2):223–233. **133**

Lewis, P. O. 1998. Maximum likelihood as an alternative to parsimony for inferring phylogeny using nucleotide sequence data. In *Molecular systematics of plants II*, edited by D. E. Soltis, P. S. Soltis, and J. J. Doyle. Boston: Kluwer. **259**

Lewis, P. O. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology* 50 (6):913–925. **259**

Lewis, P. O., M. T. Holder, and K. E. Holsinger. 2005. Polytomies and Bayesian phylogenetic inference. *Systematic Biology* 54 (2):241–253. **259**

Linder, H. P., and M. D. Crisp. 1995. *Nothofagus* and Pacific biogeography. *Cladistics* 11 (1):5–32. **381**

Liu, K., S. Raghavan, S. Nelesen, C. R. Linder, and T. Warnow. 2009. Rapid and accurate large-scale coestimation of sequence alignments and phylogenetic trees. *Science* 324 (5934):1561–1564. **207**

Liu, L., and D. K. Pearl. 2007. Species trees from gene trees: Reconstructing Bayesian posterior distributions of a species phylogeny using estimated gene tree distributions. *Systematic Biology* 56 (3):504–514. **294, 297**

Longhorn, S. J., H. W. Pohl, and A. P. Vogler. 2010. Ribosomal protein genes of holometabolan insects reject the Halteria, instead revealing a close affinity of Strepsiptera with Coleoptera. *Molecular Phylogenetics and Evolution* 55 (3):846–859. **283**

Lyell, C. 1832. *Principles of geology*. Vol. II. London: John Murray. **12**

Maddison, D. R., and W. P. Maddison. 2000. *MacClade 4: Analysis of phylogeny and character evolution*. Sunderland, Massachusetts: Sinauer. **68, 100, 315**

Maddison, D. R. 1991. The discovery and importance of multiple islands of most parsimonious trees. *Systematic Zoology* 40 (3):315–328. **207**

Maddison, W. P. 1990. A method for testing the correlated evolution of two binary characters—are gains or losses concentrated on certain branches of a phylogenetic tree? *Evolution* 44 (3):539–557. **340**

Maddison, W. P. 1997. Gene trees in species trees. *Systematic Biology* 46 (3):523–536. **166**

Maddison, W. P., M. J. Donoghue, and D. R. Maddison. 1984. Outgroup analysis and parsimony. *Systematic Zoology* 33 (1):83–103. **68**

Maddison, W. P., and L. L. Knowles. 2006. Inferring phylogeny despite incomplete lineage sorting. *Systematic Biology* 55(1): 21–30. **292, 297**

Maddison, W. P., P. E. Midford, and S. P. Otto. 2007. Estimating a binary character's effect on speciation and extinction. *Systematic Biology* 56 (5):701–710. **379, 381**

Martins, E. P., and T. F. Hansen. 1997. Phylogenies and the comparative method: A general approach to incorporating phylogenetic information into the analysis of interspecific data. *American Naturalist* 149 (4):646–667. **340**

Martinsen, E. S., S. L. Perkins, and J. J. Schall. 2008. A three-genome phylogeny of malaria parasites (*Plasmodium* and closely related genera): Evolution of life-history traits and host switches. *Molecular Phylogenetics and Evolution* 47 (1):261–273. **385**

Mau, B., M. A. Newton, and B. Larget. 1999. Bayesian phylogenetic inference via Markov chain Monte Carlo methods. *Biometrics* 55 (1):1–12. **259**

Mayden, R. L. 1999. Consilience and a hierarchy of species concepts: Advances toward closure on the species puzzle. *Journal of Nematology* 31 (2):95–116. **166**

Mayr, E. 1982. *The growth of biological thought*. Cambridge, Massachusetts: Harvard University Press. **30**

Mayr, E. 2001. *What evolution is*. New York: Basic Books. **30**

Medina, M., A. G. Collins, J. D. Silberman, and M. L. Sogin. 2001. Evaluating hypotheses of basal animal phylogeny using complete sequences of large and small subunit rRNA. *Proceedings of the National Academy of Sciences of the United States of America* 98 (17):9707–9712. **60**

Miller, J. A. 2007. Repeated evolution of male sacrifice behavior in spiders correlated with genital mutilation. *Evolution* 61 (6):1301–1315. **313–315**

Milligan, B. G. 1994. Estimating evolutionary rates for discrete characters. Pp. 299-311 In *Models in phylogeny reconstruction*, edited by R. W. Scotland, D. J. Siebert, and D. M. Williams. Oxford: Clarendon Press. **323**

Mindell, D. P., and A. Meyer. 2001. Homology evolving. *Trends in Ecology & Evolution* 16 (8):434–440. **100**

Mishler, B. D., and M. J. Donoghue. 1982. Species concepts: A case for pluralism. *Systematic Zoology* 31 (4):491–503. **166**

Morrison, D. A. 2009. A framework for phylogenetic sequence alignment. *Plant Systematics and Evolution* 282 (3–4):127–149. **207**

Morrone, J. J., and J. V. Crisci. 1995. Historical biogeography—introduction to methods. *Annual Review of Ecology and Systematics* 26:373–401. **381**

Nelson, G. J. 1971. Paraphyly and polyphyly—redefinitions. *Systematic Zoology* 20 (4):471–472. **133**

Nixon, K. C. 1999. The parsimony ratchet, a new method for rapid parsimony analysis. *Cladistics* 15 (4):407–414. **207**

Nylander, J. A. A., F. Ronquist, J. P. Huelsenbeck, and J. L. Nieves-Aldrey. 2004. Bayesian phylogenetic analysis of combined data. *Systematic Biology* 53 (1):47–67. **297**

O'Grady, P. M., R. H. Baker, C. M. Durando, W. J. Etges, and R. DeSalle. 2001. Polytene chromosomes as indicators of phylogeny in several species groups of *Drosophila*. *BMC Evolutionary Biology* 1:6. **394**

O'Hara, R. J. 1988. Homage to Clio, or, toward a historical philosophy for evolutionary biology. *Systematic Zoology* 37 (2):142–155. **306**

O'Hara, R. J. 1992. Telling the tree: Narrative representation and the study of evolutionary history. *Biology & Philosophy* 7:135–160. **30**

O'Hara, R. J. 1997. Population thinking and tree thinking in systematics. *Zoologica Scripta* 26 (4):323–329. **28, 30, 35**

Omland, K. E., L. G. Cook, and M. D. Crisp. 2008. Tree thinking for all biology: The problem with reading phylogenies as ladders of progress. *Bioessays* 30 (9):854–867. **30**

Page, R. D. M. 2003. *Tangled trees: Phylogeny, cospeciation and coevolution.* Chicago: University of Chicago Press. **381**

Page, R. D. M., and E. C. Holmes. 1998. *Molecular evolution: A phylogenetic approach.* Oxford: Blackwell. **100, 166**

Pagel, M. 1994. Detecting correlated evolution on phylogenies—a general method for the comparative analysis of discrete characters. *Proceedings of the Royal Society B-Biological Sciences* 255 (1342):37–45. **323, 340**

Panchen, A. L. 1992. *Classification, evolution and the nature of biology.* Cambridge: Cambridge University Press. **30**

Patterson, C. 1982. Morphological characters and homology. In *Problems of phylogenetic reconstruction*, edited by K. A. Joysey and A. E. Friday. New York: Academic Press. **100**

Penny, D., L. R. Foulds, and M. D. Hendy. 1982. Testing the theory of evolution by comparing phylogenetic trees constructed from five different protein sequences. *Nature* 297 (5863):197–200. **21, 23, 30**

Platnick, N. I. 1977. Paraphyletic and polyphyletic groups. *Systematic Zoology* 26 (2):195–200. **133**

Platnick, N., and G. Nelson. 1978. A method of analysis for historical biogeography. *Systematic Zoology* 27 (1):1–16. **381**

Podani, J. 2010. Monophyly and paraphyly: A discourse without end? *Taxon* 59 (4):1011–1015. **133**

Posada, D., and K. A. Crandall. 2001. Selecting the best-fit model of nucleotide substitution. *Systematic Biology* 50 (4):580–601. **259**

Poux, C., P. Chevret, D. Huchon, W. W. de Jong, and E. J. P. Douzery. 2006. Arrival and diversification of caviomorph rodents and platyrrhine primates in South America. *Systematic Biology* 55 (2):228–244. **278–279**

Procheç, S., J. R. U. Wilson, and R. M. Cowling. 2006. How much evolutionary history in a 10 × 10 m plot? *Proceedings of the Royal Society B-Biological Sciences* 273 (1590):1143–1148. **360–361**

Quicke, D. L. J., J. Taylor, and A. Purvis. 2001. Changing the landscape: A new strategy for estimating large phylogenies. *Systematic Biology* 50 (1):60–66. **207**

Rambaut, A., and N. C. Grassly. 1997. Seq-Gen: An application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Computer Applications in the Biosciences* 13 (3):235–238. **281**

Raubeson, L. A., and R. K. Jansen. 1992. Chloroplast DNA evidence on the ancient evolutionary split in vascular land plants. *Science* 255 (5052):1697–1699. **394**

Ree, R. H. 2005. Detecting the historical signature of key innovations using stochastic models of character evolution and cladogenesis. *Evolution* 59(2):257–265. **378**

Ree, R. H., and M. J. Donoghue. 1999. Inferring rates of change in flower symmetry in asterid angiosperms. *Systematic Biology* 48 (3):633–641. **321–322**

Ree, R. H., B. R. Moore, C. O. Webb, and M. J. Donoghue. 2005. A likelihood framework for inferring the evolution of geographic range on phylogenetic trees. *Evolution* 59 (11):2299–2311. **355**

Rodriguez, F., F. Wu, C. Ané, S. Tanksley, and D. M. Spooner. 2009. Do potatoes and tomatoes have a single evolutionary history, and what proportion of the genome supports that history? *BMC Evolutionary Biology* 9:191. **295**

Ronquist, F. 1997. Dispersal-vicariance analysis: A new approach to the quantification of historical biogeography. *Systematic Biology* 46 (1):195–203. **381**

Rzhetsky, A., and M. Nei. 1992. Statistical properties of the ordinary least-squares, generalized least-squares and minimum-evolution methods of phylogenetic inference. *Journal of Molecular Evolution* 35 (4):367–375. **259**

Saitou, N., and M. Nei. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4 (4):406–425. **259**

Sanderson, M. J. 1995. Objections to bootstrapping phylogenies: A critique. *Systematic Biology* 44 (3):299–320. **297**

Sanderson, M. J. 2002. Estimating absolute rates of molecular evolution and divergence times: A penalized likelihood approach. *Molecular Biology and Evolution* 19 (1):101–109. **381**

Sanderson, M. J., and M. J. Donoghue. 1994. Shifts in diversification rate with the origin of angiosperms. *Science* 264 (5165):1590–1593. **381**

Sanderson, M. J., and J. A. Doyle. 2001. Sources of error and confidence intervals in estimating the age of angiosperms from *rbc*L and 18S rDNA data. *American Journal of Botany* 88 (8):1499–1516. **381**

Sanderson, M. J., and J. Kim. 2000. Parametric phylogenetics? *Systematic Biology* 49 (4):817–829. **207**

Sanderson, M. J., and H. B. Shaffer. 2002. Troubleshooting molecular phylogenetic analyses. *Annual Review of Ecology and Systematics* 33:49–72. **297**

Sargent, R. D. 2004. Floral symmetry affects speciation rates in angiosperms. *Proceedings of the Royal Society B-Biological Sciences* 271 (1539):603–608. **388**

Schluter, D., T. Price, A. O. Mooers, and D. Ludwig. 1997. Likelihood of ancestor states in adaptive radiation. *Evolution* 51 (6):1699–1711. **318, 340**

Schrago, C. G., and C. A. M. Russo. 2003. Timing the origin of New World monkeys. *Molecular Biology and Evolution* 20 (10):1620–1625. **6**

Schwander, T., and B. J. Crespi. 2009. Twigs on the tree of life? Neutral and selective models for integrating macroevolutionary patterns with microevolutionary processes in the analysis of asexuality. *Molecular Ecology* 18 (1):28–42. **381**

Scotland, R. W. 2010. Deep homology: A view from systematics. *Bioessays* 32 (5):438–449. **100**

Shimodaira, H. 2002. An approximately unbiased test of phylogenetic tree selection. *Systematic Biology* 51 (3):492–508. **297**

Shimodaira, H., and M. Hasegawa. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution* 16 (8):1114–1116. **297**

Siddall, M. E., and M. F. Whiting. 1999. Long-branch abstractions. *Cladistics* 15 (1):9–24. **207**

Slowinski, J. B., and C. Guyer. 1993. Testing whether certain traits have caused amplified diversification—an improved method based on a model of random speciation and extinction. *American Naturalist* 142 (6):1019–1024. **381**

Sober, E. 1983. Parsimony in systematics: Philosophical issues. *Annual Review of Ecology and Systematics* 14:335–357. **100**

Sober, E. 1991. *Reconstructing the past: Parsimony, evolution and inference.* Cambridge, Massachusetts: MIT Press. **100**

Steel, M., and D. Penny. 2010. Origins of life: Common ancestry put to the test. *Nature* 465 (7295):168–169. **30**

Sullivan, J. 2005. Maximum-likelihood methods for phylogeny estimation. In *Molecular evolution: Producing the biochemical data, Part B.* San Diego: Elsevier Academic Press Inc. **297**

Swofford, D. L. 2002. *PAUP*: Phylogenetic analysis using parsimony (*and other methods). Version 4.* Sunderland, Massachusetts: Sinauer. **193, 200**

Swofford, D. L., G. J. Olsen, P. J. Waddell, and D. M. Hillis. 1996. Phylogenetic inference. In *Molecular systematics*, edited by D. M. Hillis, C. Moritz, and B. K. Mable. Sunderland, Massachusetts: Sinauer. **207, 234, 259, 297**

Teeling, E. C., M. S. Springer, O. Madsen, P. Bates, S. J. O'Brien, and W. J. Murphy. 2005. A molecular phylogeny for bats illuminates biogeography and the fossil record. *Science* 307 (5709):580–584. **99**

Templeton, A. R. 1983. Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and the apes. *Evolution* 37 (2):221–244. **273, 297**

Theobald, D. L. 2010. A formal test of the theory of universal common ancestry. *Nature* 465 (7295):219–222. **30**

Thorne, J. L., H. Kishino, and I. S. Painter. 1998. Estimating the rate of evolution of the rate of molecular evolution. *Molecular Biology and Evolution* 15 (12):1647–1657. **381**

Velasco, J. D. 2009. When monophyly is not enough: Exclusivity as the key to defining a phylogenetic species concept. *Biology & Philosophy* 24 (4):473–486. **166**

Wagner, G. P. 1989. The biological homology concept. *Annual Review of Ecology and Systematics* 20:51–69. **100**

Wake, D. B. , M. H. Wake, and C. D. Specht. 2011. Homoplasy: From detecting pattern to determining process and mechanism of evolution. *Science* 331 (6020):1032–1035. **100**

Wakeley, J. 2009. *Coalescent theory: An introduction*. Greenwood Village, Colorado: Roberts and Company. **166**

Webb, C. O. 2000. Exploring the phylogenetic structure of ecological communities: An example for rain forest trees. *American Naturalist* 156 (2):145–155. **359**

Webb, C. O., D. D. Ackerly, M. A. McPeek, and M. J. Donoghue. 2002. Phylogenies and community ecology. *Annual Review of Ecology and Systematics* 33:475–505. **381**

Weiblen, G. D., and G. L. Bush. 2002. Speciation in fig pollinators and parasites. *Molecular Ecology* 11 (8):1573–1578. **358**

Wheeler, W. 1996. Optimization alignment: The end of multiple sequence alignment in phylogenetics? *Cladistics* 12 (1):1–9. **207**

Wheeler, W. 2001. Homology and the optimization of DNA sequence data. *Cladistics* 17 (1):S3–S11. **207**

White, M. A., C. Ané, C. N. Dewey, B. R. Larget, and B. A. Payseur. 2009. Fine-scale phylogenetic discordance across the house mouse genome. *PLoS Genetics* 5 (11): e1000729. **145–146**

Whiting, M. F., J. C. Carpenter, Q. D. Wheeler, and W. C. Wheeler. 1997. The Strepsiptera problem: Phylogeny of the holometabolous insect orders inferred from 18S and 28S ribosomal DNA sequences and morphology. *Systematic Biology* 46 (1):1–68. **282–283**

Wiley, E. O. 1981. *Phylogenetics: The theory and practice of phylogenetic systematics*. New York: Wiley. **133, 207**

Wiley, E. O., D. R. Brooks, D. Siegel-Causey, and V. A. Funk. 1991. *The Compleat Cladist: A Primer of Phylogenetic Procedures*. Lawrence, Kansas: Kansas Museum of Natural History. **207**

Wilkinson, M., P. R. Peres-Neto, P. G. Foster, and C. B. Moncrieff. 2002. Type 1 error rates of the parsimony permutation tail probability test. *Systematic Biology* 51 (3):524–527. **297**

Wyss, A. R., and J. J. Flynn. 1993. A phylogenetic analysis and definition of the Carnivora. In *Mammal phylogeny: Placentals*, edited by F. S. Szalay, M. J. Novacek, and M. C. McKenna. New York: Springer-Verlag. **175, 181, 193**

Yoder, A. D., M. M. Burns, S. Zehr, T. Delefosse, G. Veron, S. M. Goodman, and J. J. Flynn. 2003. Single origin of Malagasy Carnivora from an African ancestor. *Nature* 421 (6924):734–737. **369–370**

Zanis, M. J., D. E. Soltis, P. S. Soltis, S. Mathews, and M. J. Donoghue. 2002. The root of the angiosperms revisited. *Proceedings of the National Academy of Sciences of the United States of America* 99 (10):6848–6853. **303**

Zimmer, C. 2009. *The tangled bank: An introduction to evolution*. Greenwood Village, Colorado: Roberts and Company. **30**

Glossary

absolute chronogram A tree in which the ages of nodes and/or the actual duration of branches is shown.

acyclic A graph in which there is one and only one path between any two points; a tree is an acyclic graph.

adaptation A derived feature that went to fixation because it was favored by selection (relative to the ancestral state) in the lineage in which it evolved. It is sometimes also specified that adaptations continue to serve the function for which they initially evolved.

additivity A property of a tree (or part of a tree) in which the sum of the branch lengths between each pair of taxa is equal to their pairwise distance.

alignment (= sequence alignment) The process of adding gaps in a matrix of DNA sequences to line up homologous nucleotide positions, or the product of this process.

allele (= haplotype) A copy of a gene found at a particular locus, which may or may not differ in sequence from the alleles found in other individuals in the population.

allopatric Occurring in separate geographic areas.

ancestor An individual or population from which more recent individuals or populations have descended. In the context of phylogenetic trees, branches and nodes represent populations of ancestors of the taxa at the tips of the tree.

ancestral A character state, allele, or node that was present before the evolution of a derived character state, allele, or node. The term should not be applied to living taxa.

ancestral character state The character state inferred to have been present in a particular ancestor.

apomorphy A derived character state (opposite = plesiomorphy).

autapomorphy A derived character state that is restricted to one taxon in a particular data set.

autocorrelated relaxed molecular clock See correlated relaxed molecular clock.

Bayesian Statistical approaches that formally incorporate prior knowledge or belief in determining the posterior probability that a hypothesis is true.

Bayesian majority-rule consensus tree (= Bayesian tree) A consensus tree composed of clades with a posterior probability higher than some threshold (usually 0.5) that is used to summarize the results of a Bayesian MCMC phylogenetic analysis.

Bayesian tree See Bayesian majority-rule consensus tree.

binary When applied to a variable or a character: existing in two states. When applied to a tree diagram: with all nodes yielding two descendant lineages (dichotomous).

binary-state speciation and extinction (BiSSE) A model in which a single character can evolve between two states that have potentially different effects on the rate of lineage splitting and/or extinction.

binomial The naming of a species in traditional taxonomy through the use of two names, a genus name and a specific epithet, e.g., *Homo sapiens*.

biogeography The field of study focused on understanding biological distribution patterns and how they have changed over time.

biological species concept A concept that defines species as groups of potentially interbreeding organisms that are reproductively isolated from other biological species.

bipartitions See splits.

bootstrap See bootstrap percentage or bootstrap analysis.

bootstrap analysis (= bootstrap) A method for assessing confidence in a statistical inference by generating many random, pseudoreplicate data sets and analyzing each pseudoreplicate to see how frequently a particular result is obtained. Includes nonparametric bootstrap and parametric bootstrap.

bootstrap consensus tree (= bootstrap tree) A consensus tree that summarizes clades found in analyses of bootstrap replicates. Most commonly, a 50% majority-rule consensus (See majority-rule consensus tree).

bootstrap percentage (= bootstrap score; bootstrap) The percentage of bootstrap replicates whose optimal trees include a particular clade, which is used as a measure of support for that clade.

bootstrap sample See pseudoreplicate.

bootstrap score See bootstrap percentage.

bootstrap tree See bootstrap consensus tree.

branch (= edge) The portion of a phylogenetic tree that connects two nodes (internal branch) or one node and a tip (external branch).

branch-based definition A clade definition referring to the largest clade that includes the internal specifiers and excludes the external specifiers.

Brownian motion A random walk, originally used to describe the movement of small particles in liquid, which provides a useful model for the evolution of continuous traits over time, where values fluctuate up and down stochastically but the expected net change is zero (i.e., change is not directional).

burn-in A set of samples from the beginning of the Markov chain in Bayesian phylogenetics, before the chain has reached stationarity (i.e., before the chain has found

a region of parameter space with high posterior probability). Because of sensitivity to starting parameter values, the burn-in provides a poor estimate of the posterior distribution, which is why the burn-in samples are usually discarded.

calibration The use of fossils or other pieces of information to help convert a tree's branch lengths to units of absolute time.

character A heritable feature of a taxon that can be used to infer the phylogeny or whose evolution can be studied using phylogenetic trees built from other sources of information.

character encoding The process of dividing the observed variation in a morphotypic character into discrete characters and character states.

character polarity The direction of character state change. The specification of which character states are ancestral and which are derived.

character scoring The process of assigning character states to taxa for the purpose of phylogenetic inference or phylogenetic comparative methods.

character state A specific form of a character that occurs in some taxa.

character state–dependent diversification The effect of alternative character states on rates of lineage splitting (birth) and/or extinction (death). See also binary-state speciation and extinction.

character state matrix A table showing the state of each character occurring in each taxon. Generally, rows represent taxa, columns represent characters, and numbers or letters in the cells represent the character states.

character state weighting An application of generalized parsimony that assigns some character state transitions greater weight than others.

chronogram A tree in which branch lengths are proportional to time. May either be an absolute chronogram or a relative chronogram.

circle tree A tree-drawing format in which the branches are curved and bent so that the root is in the middle and the tips are arranged in a circle.

clade A monophyletic group. A part of a rooted tree that can be separated from the rest of the tree (which includes the root) by cutting a single branch.

clade credibility The posterior probability that a clade is true in a Bayesian phylogenetic framework.

cladogenesis See lineage splitting.

cladogram A tree diagram that communicates the topology (the pattern of branching) but not the branch lengths. A cladogram provides all the information relevant to determining the degree of evolutionary relatedness between taxa.

clonal Reproducing asexually.

coalescence Looking backward in time, the point at which two alleles (gene lineages) converged on a single common ancestor.

coalescence theory A branch of population genetics that models the expected pattern of coalescence in branching and unbranching population lineages.

coalescence units The product of the population size and the number of generations. A measure of branch length used to predict the probability of deep coalescence.

codes of nomenclature Sets of rules that govern how species and other taxa are named.

codominant trees Two (or more) gene trees that each apply to a significant proportion of the genome and occur at roughly equal frequency.

combined analysis (= concatenation) When two or more data partitions (e.g., two genes or a gene plus a set of morphological characters) are concatenated into a single data set before phylogenetic analysis.

common ancestry The existence of ancestral organisms or taxa that are ancestral to more than one taxon (living or extinct).

comparative methods Studies of the distribution of trait variation among species, which often focus on evaluating the role of natural selection during trait evolution. Modern comparative methods are statistically rigorous and take account of phylogenetic relationships. Sometimes also referred to as *the* comparative method.

compatible Describes a tree that has the same topology as a second tree for the subset of taxa that are shared between the two trees. Also describes a character whose evolution can be explained on a specified tree without invoking homoplasy; a consistent character.

concatenation See combined analysis.

concordance factor The proportion of genes in a genome whose gene trees have a given clade.

concordance tree A tree composed of those clades that are estimated to be shared by a plurality of the gene trees in the genome.

conflict When two or more characters or whole data sets support different trees (i.e., trees with incompatible clades).

consensus tree A tree that summarizes the clades or relationships that are shared among a group of input trees.

consistency The degree to which a character can be reconciled with a tree without invoking homoplasy.

consistency index A measure of how well or poorly a particular character or set of characters fits a tree: the minimum number of steps possible divided by the number

of steps required to explain the character(s) on the tree. Trees with higher consistency indices have less homoplasy.

consistent A character that can be reconciled with a tree without invoking any homoplasy, that is, by invoking only the minimum number of changes. Also referred to as a compatible character.

constrained search A search that is restricted to tree topologies that have a specified topological feature (e.g., having or lacking a specified clade).

constraint A topology that is used to limit a phylogenetic analysis to a subset of tree space.

constraint tree The optimal tree found during a constrained search. Sometimes also used for the unresolved tree that is used to communicate a topological constraint to a computer program.

continuous Used to describe parameters (e.g., substitution rates) or traits (e.g., body size) that can take any value along a continuum.

contrast A comparison of the trait values in sister groups (e.g., used in phylogenetic independent contrasts).

convergent evolution An evolutionary pattern in which the same character state arises independently in different lineages.

cophylogenetics Phylogenetic studies of symbiotic taxa that might have affected one another's lineage-splitting history; often undertaken with the goal of testing for cospeciation.

correlated evolution A pattern in which two or more discrete or continuous characters affect each other's evolution—the state of one trait affects the rate of change in the other trait. Correlated evolution can be used as a source of evidence of adaptation.

correlated relaxed molecular clock (= autocorrelated relaxed molecular clock) A model in which different branches are permitted to have different rates of molecular evolution where lineages tend to have a rate that is correlated with their immediate ancestral and descendant lineages.

cospeciation When symbiotic taxa show correlated lineage splitting as a result of their close ecological interactions.

cost (= parsimony cost) In generalized parsimony: the product of a character's length (number of steps) multiplied by the character's weight.

crown group A clade including all the tips, the extinct branches, and the common ancestor, but not the branch subtending that ancestor (the stem lineage).

crown node The node that represents the last common ancestor of a clade of interest.

data matrix See character state matrix.

decay index A parsimony-based measure of support for a clade based on the difference in tree length between the shortest tree with the clade and the shortest tree that lacks the clade.

deep coalescence When the coalescence of two gene lineages occurs farther back in time than the point at which the populations or species containing those lineages diverged.

deletion When nucleotides are removed from a DNA sequence during its evolution.

depth See nodal depth.

derived A character state or allele that arose during evolution and replaced an ancestral character state or allele. The term should not be applied to taxa.

diagnostic traits Traits that are used to distinguish one species from another.

diagonal tree A tree diagram format in which the nodes are connected by straight diagonal lines.

dichotomous In relation to tree diagrams: a pattern in which an ancestral lineage gives rise to exactly two descendant branches. See also binary.

directed A graph in which each edge has an implicit directionality. A rooted tree (but not an unrooted tree) is a directed graph, where the direction is provided by the time axis.

discordant In reference to gene trees, when gene trees disagree (= conflict) with each other.

dispersal The movement of individuals or populations to a new geographic location.

diversification rate The rate at which the number of lineages in a clade grows; usually equal to the rate of lineage splitting minus the rate of extinction.

dominant history A single gene tree that applies to the majority of the genome even in cases where there is genealogical discordance.

edge See branch.

equally weighted parsimony (= flat-weighted parsimony) A form of parsimony in which all characters and character state changes have the same cost.

evolutionary correlation See correlated evolution.

evolutionary distance The length of the path along the tree that separates two taxa or nodes measured in units of the expected number of substitutions per site.

exaptation A trait of a particular taxon that originally evolved for one function but was later co-opted for a different function that it serves in the taxon of interest.

exclusivity A property of a taxon whose member organisms are more closely related to one another than to any organisms outside the taxon.

external branch See terminal branch.

external specifier In relation to phylogenetic nomenclature: a tip that is outside of a clade, but is used to attach a name to that clade as part of a branch-based definition.

Fitch parsimony See equally weighted parsimony.

fixed An allele or character state that has a frequency of 1.0, meaning that it is carried by all individuals in the population.

flat prior probability (= uninformative prior probability) A prior probability that is set to be equal over a range of parameter values. Alternate tree topologies are often assigned flat prior probabilities. Also referred to as a uniform prior probability.

flat-weighted parsimony See equally weighted parsimony.

fully resolved Referring to a tree that lacks polytomies.

gap cost See gap penalty.

gap penalty (= gap cost) A parameter used during multiple alignment to penalize alignments that invoke indels; typically, a gap penalty is expressed in units of the number of substitutions that need to be avoided for an indel to be invoked.

gaps Positions in a matrix of DNA sequences that indicate that an insertion or deletion event (= indel) has occurred.

gene family A group of genes that have arisen by gene duplication from an ancestral gene copy.

gene flow The movement of genes between populations that links individual populations together into evolutionary lineages.

gene tree The history of a set of homologous recombinational genes.

genealogical concordance Agreement among gene trees due to the genes having tracked the same population history.

genealogical discordance Disagreement among gene trees due to incomplete lineage sorting or reticulation.

generalized parsimony A form of parsimony in which characters and character state changes may be weighted differently based on prior knowledge about how they evolve, and where the best hypothesis is the one that minimizes the weighted sum of all changes (the number of changes of each type multiplied by their respective weights).

genetic drift A chance change in allele frequency, due to random sampling in a finite population.

genus A taxonomic rank above species.

grade See paraphyletic group.

haplotype See allele.

hard polytomy A polytomy in a tree diagram that is caused by the simultaneous origin of three or more descendant lineages from the same ancestral lineage.

heritability The tendency for ancestors and their descendants to manifest similar traits.

heuristic search A search that is designed to find optimal trees in tree space without necessarily considering all possible trees.

hierarchical likelihood ratio test (= likelihood ratio test) A way to use likelihood scores to choose among competing, nested probabilistic models. Model fit is usually judged using a chi-square statistic, which is equal to twice the difference in log-likelihood, with the degrees of freedom equal to the number of extra parameters in the more complex model.

historical chronicle A description of a series of events in the order in which they occurred.

historical narrative A description of a series of events, the reason why they occurred, and/or the cause-and-effect connections between events.

homologous genes (= homologs) Gene sequences that were derived from the same common ancestral sequence. Includes orthologs and paralogs.

homologous traits Traits in two organisms or taxa that were inherited from the common ancestor of the two organisms/taxa.

homologs See homologous genes.

homology The sharing of features due to common ancestry.

homoplasy A lack of consistency between a tree topology and the distribution of character states among taxa. Includes convergent evolution and reversal.

horizontal gene transfer See lateral gene transfer.

hybrid speciation When members of two distinct lineages interbreed and form a new evolutionary lineage that is distinct from the two parental lineages (if they persist). See also lineage fusion.

incomplete lineage sorting The maintenance of genetic variation within a population lineage from one population-splitting event to the next such that deep coalescence can occur; a common cause of genealogical discordance.

incongruence length difference test (= partition homogeneity test) A test used to assess whether two data partitions have conflicting phylogenetic signals.

indel An insertion or deletion event in a DNA sequence.

inference A conclusion reached based on observations combined with background assumptions.

ingroup A group of taxa under study that is assumed to constitute a clade.

input tree One of multiple trees being summarized in a consensus tree method.

insertion The addition of one or a stretch of nucleotides in a DNA sequence.

instantaneous rate The expected rate at which substitutions or changes of character state occur.

internal branch (= internode) The branches of a phylogenetic tree that connect nodes.

internal specifier In relation to phylogenetic nomenclature: a tip that is inside a clade and is used to attach a name to that clade as part of a branch-based or node-based clade definition.

internode See internal branch.

introgression When individuals from two distinct lineages produce hybrid offspring that subsequently interbreed with one of the parental lineages and cause genes from one lineage to be introduced into (and maybe become fixed in) the other lineage. A source of discordance between gene trees.

key innovation A trait that increases the diversification rate of a lineage.

ladder thinking Viewing evolution incorrectly as a ladder of progress with some species as primitive (near the bottom of the ladder) and others as advanced (near the top).

Lamarckianism A theory of biological evolution proposed by Jean Baptiste Lamarck in which life arises by spontaneous generation and evolves into more advanced forms by the use and disuse of parts.

lateral gene transfer (= horizontal gene transfer) When a small piece of DNA moves between two lineages by a process other than sexual reproduction. A source of discordance between gene trees.

leaf See tip.

length See tree length.

likelihood The probability of the data under a certain hypothesis (e.g., a tree topology). Calculation of the likelihood requires the specification of a probabilistic model.

likelihood ratio A comparison of the likelihoods of two hypotheses obtained by dividing one likelihood by the other or by subtracting one log-likelihood from the other.

likelihood ratio test See hierarchical likelihood ratio test

lineage A group of populations united by occasional gene flow. Sometimes also used (informally) to refer to a clade.

lineage birth rate The rate at which new lineages arise by splitting of an existing lineage.

lineage death rate The rate at which lineages go extinct.

lineage fusion When two distinct evolutionary lineages merge into a single, hybrid descendant lineage. See also hybrid speciation.

lineage sorting The process by which alleles (gene lineages) that formally coexisted in a population are lost over the course of evolution.

lineage splitting The division of an ancestral lineage into two or more descendant lineages.

locus A region of the genome that may exist as different alleles within a population.

log-likelihood The natural logarithm of the likelihood.

long-branch attraction A phylogenetic artifact, arising when rates of evolution are high on two or more unrelated lineages, that tends to result in descendants of these "long branches" being united erroneously into the same clade.

majority-rule consensus tree A consensus tree that includes only clades present in greater than some specified percentage of the input trees (usually 50%).

mapping Reconstructing character evolution on a tree, usually by flat-weighted parsimony.

Markov chain Monte Carlo (= MCMC) A computational method used in a Bayesian framework for estimating the posterior distribution of trees and other parameters.

Markov model A model in which the probability of a change is dependent only on the current state, not on the path that led to the current state.

maximum likelihood Statistical method in which the best estimate of a parameter (e.g., a tree) is that which has the highest probability of yielding the observed data (e.g., DNA sequences) while picking the values of all nuisance parameters that maximize the likelihood.

maximum parsimony In phylogenetics, a method in which the best hypothesis is taken to be that which minimizes the total cost of the implied character state changes. Includes equally weighted parsimony and generalized parsimony.

MCMC See Markov chain Monte Carlo.

MCMC generation A subset of a Markov chain Monte Carlo run that involves the proposal of a new parameter combination followed by the acceptance or rejection of this proposal.

merging A maneuver for simplifying a tree by representing a clade with a single, appropriately labeled tip.

minimizing deep coalescences A method for inferring a species tree from multiple gene trees by choosing the population tree that requires the fewest deep coalescence events to explain all of the gene trees.

minimum evolution A distance-based method that, for each topology, finds the set of branch lengths that maximize the fit between observed and expected pairwise distances, and then selects the topology that has the smallest summed branch length.

minor tree A gene tree that differs from the dominant history. Can result from processes such as incomplete lineage sorting or lateral gene transfer. Also called a minor history.

mixing Refers to the exploration of parameter space in a Bayesian MCMC analysis; in an analysis with adequate mixing, the post–burn-in samples should approximate the true posterior distribution.

model-based Describing an inference method that uses a mathematical model to calculate the probability of particular outcomes.

molecular clock A model of molecular evolution in which the rate of evolution is the same on all branches of the tree such that branch lengths are directly proportional to time.

molecular dating The process of estimating the ages of nodes in a phylogeny based on statistical analyses of molecular data and the fossil record.

monophyletic Composed of an ancestor and all of its descendants (i.e., being a clade).

monotypic A ranked taxon that includes only one taxon at a lower rank (e.g., a family composed of only one genus).

morphology The physical features of an organism.

morphotypic A general term for characters such as morphology, which exist in discrete states where the universe of possible character states is not constrained in advance but is inferred by observing the states seen among a set of taxa.

most parsimonious A tree or a mapping of characters on a tree that explains a data set with the lowest parsimony cost.

multifurcation See polytomy.

mutation A substitution in a DNA sequence, which may or may not result in a change in the organism's phenotype.

natural selection The differential survival and reproduction of individuals due to their inherited characteristics; a cause of evolution that can rapidly increase an allele's frequency in a population.

neighbor-joining An algorithm that rapidly builds a tree from a distance matrix.

nested In describing clades on a tree: when a smaller clade is part of a larger clade. In describing likelihood models: when a simpler model is a special case of a more complex model, meaning that the simpler (nested) model has only a subset of the free parameters of the more complex model.

nodal (or node) depth (= nodal height) The evolutionary distance from a particular node to the tips of an ultrametric tree.

nodal height See nodal depth.

node A branching point in a phylogenetic tree.

nomenclature A system regulating the application of names to taxa.

nonparametric Methods that draw inferences without assuming an explicit probabilistic model.

nonparametric bootstrap A bootstrap analysis in which the replicate data sets are generated from the original data set by sampling with replacement.

nuisance parameter A parameter that needs to be specified before the likelihood of a hypothesis can be calculated but that is not a parameter whose value we actually care to estimate. Maximum likelihood and Bayesian approaches differ in how they typically handle nuisance parameters.

observations Facts we accept to be true when making statistical inferences.

optimality criterion A metric of quality (e.g., length, likelihood) that is applied to trees and may be used as the basis for selecting the optimal tree for a given data set.

optimization The process of finding the best values for a particular parameter, such as branch lengths or tree topology.

Ornstein-Uhlenbeck model A stochastic model describing the evolution of a continuous trait that allows trait values to evolve randomly while being pulled toward an optimal value by stabilizing selection.

orthologous genes (= orthologs) Homologous gene sequences in different taxa that have arisen due to population lineage splitting (speciation) as opposed to gene duplication.

orthologs See orthologous genes.

outgroup Taxon or taxa that are assumed *a priori* not to be within the ingroup and may therefore be used to root the tree and polarize character states.

pairwise distance A measure of the degree of difference between two tips, for example, the proportion of characters for which the two taxa have a different character state.

parallel evolution See convergent evolution (some authors distinguish parallel and convergent evolution, but these distinctions are not relevant here). Also referred to as parallelism.

paralogous genes (= paralogs) Homologous gene sequences in different taxa whose history entails at least one duplication since their common ancestral sequence.

paralogs See paralogous genes.

parametric Methods that draw inferences by using explicit probabilistic models.

parametric bootstrap In phylogenetics: a bootstrap analysis in which the replicate data sets are generated by simulating evolution up the optimal tree obtained from the original data set.

paraphyletic group (= grade) A kind of non-monophyletic group that contains the common ancestor but not all descendants of that ancestor; characterized by one or more shared ancestral character states (plesiomorphies).

parsimony See maximum parsimony.

parsimony cost See cost.

parsimony-informative Characters whose length differs across tree topologies so that the character potentially influences which trees are judged optimal by maximum parsimony.

parsimony-uninformative Characters whose length is the same on all possible tree topologies, which therefore have no influence on which trees are judged optimal by maximum parsimony.

partition A subset of a data set composed of characters that might differ in history (tree topology and/or model of evolution) from other partitions.

partition homogeneity test See incongruence length difference test.

permutation When the character states of each character are independently shuffled among taxa so as to erase any phylogenetic signal that may have been present.

permutation tail probability test (= PTP test) A test for phylogenetic signal in a data set that compares the length of the optimal tree for the original data set to the optimal trees from permuted data sets.

phenetic species concept A concept that defines species as groups of organisms that share greater overall similarity with one another than with organisms outside the group.

phenotype An observable feature of an organism.

PhyloCode A code of phylogenetic nomenclature.

phylogenetic clustering A pattern in which the species co-occurring in a community are more closely related to one another than expected by chance (opposite = phylogenetic overdispersion).

phylogenetic community ecology A field concerned with examining the influence of phylogenetic history on the assembly of ecological communities.

phylogenetic generalized least squares A phylogenetic comparative method for continuous traits that is more flexible than phylogenetic independent contrasts, for example, in being able to accommodate the Ornstein-Uhlenbeck model of trait evolution.

phylogenetic independent contrasts A phylogenetic comparative method that assesses whether two continuous traits have evolved independently by comparing the trait values between sister groups (where some sister groups correspond to internal nodes whose trait values are not observed but are estimated from tip data).

phylogenetic inference Estimation of a phylogenetic tree based on a data set and an optimality criterion.

phylogenetic niche conservatism The tendency for lineages to remain associated with the same ecological niche through lineage-splitting events.

phylogenetic nomenclature A system of nomenclature that is designed to attach names to clades without reference to ranks.

phylogenetic overdispersion A pattern in which the species co-occurring in a community are more distantly related to one another than expected by chance (opposite = phylogenetic clustering). Also referred to as phylogenetic evenness.

phylogenetic species concept A concept that defines species as groups of organisms that are more closely related to one another than to any organisms outside the group.

phylogenetic systematics A philosophy of systematics that holds that the observations of trait variation among organisms can be used to infer the relatedness of organisms and that this inferred phylogeny should form the basis of classification, with only monophyletic groups being valid.

phylogenetics The field of study concerned with inferring the evolutionary relationships of living and extinct taxa and using this information to learn about patterns and processes of evolution.

phylogeny The evolutionary history of a group of organisms that has a tree form. A tree diagram summarizing the inferred evolutionary relationships of the included tips. Also referred to as a phylogenetic tree.

phylogram A tree diagram in which branch lengths are drawn proportional to the amount of change attributed to the branch.

plesiomorphy An ancestral character state (opposite = apomorphy).

point estimate The best estimate of a parameter given a set of data.

polarity See character polarity.

polymorphic Having two or more states within a single population or taxon.

polyphyletic group A kind of non-monophyletic group in which the group does not include the most recent common ancestor of all members of the group; characterized by a convergently evolved trait.

polyploidy The possession of more than two sets of chromosomes in a single genome.

polytomy (= multifurcation) In a rooted tree, a node with more than two descendants. In an unrooted tree, a node connected to more than three branches. Divided into hard and soft polytomies.

posterior probability The probability of a hypothesis after observing data. Equal to the likelihood of the data given the hypothesis, multiplied by the prior probability of the hypothesis, divided by the probability of the data.

posterior probability distribution (= posterior distribution) The posterior probabilities for different values of a parameter of interest.

principle of parsimony In general: the principle that posits that the best explanation of a phenomenon is the one requiring the fewest hypotheses or events. In relation to phylogenetic inference: the principle that the tree that invokes the least homoplasy, perhaps weighted to reflect prior expectations of homoplasy (in generalized parsimony), is the best estimate of the true tree.

prior probability The probability of a hypothesis based only on background knowledge, before any data are collected.

priority at rank A taxonomic rule, which states that the correct ranked name for a taxon is the earliest name published at that rank.

pruning The process of removing tips or clades from a phylogenetic tree without changing the implied relationships of the remaining tips.

pseudoreplicate (= bootstrap sample) A data set generated by bootstrap analysis on the original data, which is expected to resemble new data that could potentially be obtained.

PTP test See permutation tail probability test.

purines One of the two classes of nucleotides, composed of the bases adenine (A) and guanine (G).

pyrimidines One of the two classes of nucleotides, composed of the bases cytosine (C), thymine (T), and uracil (U).

rank A level in the traditional hierarchical system of nomenclature including, for example, species, genus, and family.

recombinational gene A contiguous segment of DNA that shares a single gene tree and is bounded by relevant recombination events that result in adjacent DNA sequences having distinct gene trees.

rectangular tree A tree composed of only vertical and horizontal lines in which the nodes are connected by lines bent at right angles.

relatedness The recency of common ancestry, with more closely related organisms being those that share more recent common ancestors.

relative chronogram A tree in which the branch lengths are directly proportional to time, but absolute time is not specified.

relaxed clock See relaxed molecular clock

relaxed molecular clock (= relaxed clock) A molecular dating approach that allows the rate of molecular evolution to vary from branch to branch within some constraints. Divided into uncorrelated and correlated relaxed molecular clocks.

reproductive isolation A lack of gene flow between populations such that they are expected to evolve independently.

resolved A portion of the tree that is fully binary, that is, containing no polytomies.

reticulate Having a netlike population history with both lineage splitting and fusion.

reversal A kind of homoplasy in which a taxon evolves a character state that had already been visited by an ancestor of the taxon.

root The earliest node in a tree, the point from which the descendant lineages arise.

rooted tree A tree with a specified time axis with the earliest (oldest) node in the tree indicated by the root.

rooting The process of identifying the root of a tree, commonly achieved by including outgroup taxa in phylogenetic analyses.

sampling with replacement The process of randomly sampling characters from an original data set while allowing that sampling a character once does not change the probability of it being sampled again; used to create replicate data sets for nonparametric bootstrap analysis.

segmental duplication The duplication of a portion of a chromosome that can result in paralogous copies for one or more genes.

selective regime The environmental conditions (including climate, resources, predators, competitors, etc.) that collectively determine the relative fitness of individuals with different traits in a population.

sequence alignment See alignment.

shortest In parsimony, describes the tree requiring the fewest changes in character state, that is, requiring the least homoplasy.

silent substitutions Mutations in DNA that do not change the phenotype. See also synonymous substitutions.

sister groups (= sister taxa) Two lineages or taxa descended from a common ancestor.

sister taxa See sister groups.

site likelihood The likelihood associated with a single character, which is obtained by determining the probability that the observed character state pattern would have evolved given a set of model parameters.

soft polytomy A polytomy in a tree diagram that reflects uncertainty as to the true order in which three or more lineages branched from one another.

species tree The tree like history of population branching. Also, the relationships among species when contrasted with the relationships among genes in those species (gene trees).

species tree inference The use of data from multiple gene trees to estimate the branching history of populations.

specific epithet The adjectival portion of a binomial species name that indicates which member of a genus is being referred to. In *Homo sapiens*, "*sapiens*" is the specific epithet.

splits (= bipartitions) Divisions of a rooted or unrooted tree into two groups of taxa.

SPR See subtree pruning and regrafting.

squared-change parsimony A method for reconstructing the evolution of a continuous trait by minimizing the square of the amount of change attributed to each branch of a phylogeny.

state See character state.

stationarity In relation to Bayesian MCMC analysis: when a chain ceases to show a trend toward increasing likelihood.

statistical power The probability that a statistical test will reject the null hypothesis when it is false.

stem lineage The branch that subtends a clade, usually used in reference to a lineage that gave rise to a noteworthy clade.

stem node The node that represents the last common ancestor shared by a taxon of interest and its closest relative.

steps Changes in character state along a tree as determined using parsimony.

stochastic mapping A Bayesian method that generates a sample of evolutionary histories for each character in such a way that the frequency of events in the sample should approximate their posterior probabilities.

strict consensus tree A consensus tree that includes only clades present in all input trees.

substitution A change in a DNA sequence in which one nucleotide is replaced with a different base.

substitution model A probabilistic model describing the categories of substitutions that can occur and the potential differences between these substitution categories in their rate of evolution.

substitution probability matrix (= transition matrix) A square matrix that contains the probabilities of moving between pairs of character states given a certain amount of time.

subtree pruning and regrafting (= SPR) A tree rearrangement algorithm that entails separating a subtree (a clade if the tree is rooted) from the main tree (the main tree being defined as the part that contains the root or a designated reference taxon) and then reattaching the subtree to a new location on the main tree.

sympatric Occurring in the same geographic area.

symplesiomorphy A shared primitive character. Taxa diagnosed only by symplesiomorphies are likely to be paraphyletic.

synapomorphy A shared derived state; taxa diagnosed by synapomorphies should be monophyletic.

synonymous substitutions Mutations in DNA that do not change the encoded amino acid sequence.

systematics The study of the evolutionary relationships among organisms (phylogeny) and the application of this knowledge to create systems of classification.

taxon (plural = taxa) A formally named group of organisms. The groups of organisms represented by the tips of a phylogenetic tree.

taxonomy The formal naming and classification of organisms.

Templeton test A non-parametric topology test that uses the magnitude of differences in length for characters across trees to assess whether a data set is significantly more compatible with one tree than with another.

terminal See tip.

terminal branch A branch that connects a node to a tip.

time-reversible A property of processes that behave in the same way whether moving forward or backward in time. In the context of substitution probability matrices, this property means that, for example, transitions from A to T will have the same probability as transitions from T to A.

tip (= terminal or leaf) The entities (e.g., taxa, genes) whose relationships are studied and/or depicted using a tree diagram.

tip correlation A correlation between traits that is detected when considering the trait values of tips without taking account of their phylogenetic relationships.

topology The history of lineage splitting depicted by a tree.

topology test A class of tests that evaluate specific phylogenetic hypotheses (e.g., the presence of a particular clade) by comparing the scores of the optimal trees that are consistent with the hypothesis with the optimal trees that contradict the hypothesis.

trait A character or character state.

trait-based species concepts A class of species concepts that define species as groups of organisms that share a trait or set of traits, such as morphological features, ecological specialization, or reproductive compatibility.

transition A DNA substitution of a purine for a purine (A to G or G to A) or a pyrimidine for a pyrimidine (C to T/U or T/U to C).

transition matrix See substitution probability matrix.

transition:transversion bias The tendency for transitions to occur at a higher frequency than transversions.

transitional fossils Fossils that possess some but not all of the traits of a particular clade and can be used to infer the order in which the unique traits of that clade evolved.

transversion A DNA substitution of a purine (A, G) for a pyrimidine (C, T, U) or vice versa.

tree An acyclic graph. May be rooted or unrooted.

tree balance The degree to which sister groups contain the same number of tips.

tree length (= length) In a parsimony framework, a count of the number of character state changes (steps) that a given tree requires to account for character variation seen among its tips.

tree score A measure applied to a tree that may be used to select among trees in a phylogenetic analysis (e.g., length, likelihood).

tree space A multidimensional space containing all possible trees for a particular set of taxa where more similar trees are closer to each other than more different ones; sometimes visualized with one additional dimension representing the tree score (length, likelihood, posterior probability). A two-dimensional representation of true tree space used to conceptualize the process of searching for optimal trees.

tree thinking The ability to read and interpret phylogenetic trees and use trees to accurately represent the evolutionary process.

tree-to-tree distance A measure of the topological differences between two trees containing the same set of tips.

trend A pattern of increasing or decreasing frequency of a character state, or a directional change in a trait's value resulting from inequalities in the rate of evolution in different directions.

type A reference specimen used to anchor a taxonomic name to a ranked taxon in traditional nomenclature.

ultrametric A rooted tree in which units of branch length accumulate evenly on all branches so that all contemporaneous tips are equidistant from the root node.

uncorrelated relaxed molecular clock A model in which different branches are permitted to have different rates of molecular evolution, with each branch having a rate that is drawn from the same parametric distribution.

uninformative prior probability See flat prior probability.

unrooted tree A tree without a root and, thus, without a specified time axis.

vestigial trait Structures that are reduced and nonfunctional in one taxon, but are present and functional in related taxa.

vicariance The division of an ancestral population into two or more descendant populations by the appearance of geographic barriers that break up the ancestral range.

whole genome duplication The doubling of an entire genome, resulting in two paralogous copies of every gene (i.e., polyploidy).

Index

absolute chronogram, 362, 369, **439**
acceptance ratio, 262
actinomorphy, 321–322, 388
acyclic graph, 43, **439**
adaptation
  compared to key innovation, 376
  cross-species comparisons, 311–312, 325, 328, 344, 347, 421
  definition of, 306, **439**
  in community assembly, 362
  in evolutionary narratives, 306, 421
  testing, 306–311, 340, 345, 421
adaptive zone, 376
adjacent trees, in tree space, 190–191
advanced taxa or traits, 10, 23, 101
AFLPs, 231, 392
Africa, 351–356
Afrotheria, 51
algorithms
  Bayesian MCMC, 252
  branch-and-bound, 189, 193
  contrast to optimality, 179
  heuristic search, 190–191
  maximum likelihood, 242
  multiple alignment, 199–200
  neighbor-joining, 234, 236
alignment. *See* sequence alignment
alleles
  coalescence, 142–150
  definitions of, 140–141, **439**
  fixation of, 78–80, 84
  introgression of, 158
  polymorphic, 77–80, 140–143, 166
  relatedness of, 150
  tips of gene trees, 141–143, 146–148, 166
allopatric speciation
  definition of, 39–40
  effect on diversification, 376
  effect on phylogenetic overdispersion, 362
altitude
  cause of vicariance, 383
  in tree space, 190, 192, 250
Amborella, 303
amino acid substitutions
  branch length, 56
  changes on tree, 101
  model of, 230, 390
amnion, 102
amniotes
  evolution in, 116–117
  taxonomy, 117–118, 120–124
amphibians, 11, 28–29, 121–124, 136
amplified fragment length polymorphisms, 231, 392
analogous traits, 3
  See also homology
ancestor-descendant relationships, 24, 35
ancestral allele, 77–80, 141–142, 148–150
ancestral character state
  definition, 439
  in likelihood calculation, 243
  in tests of adaptation, 308
  inference of, 41
  molecular evolution, 88
  resulting in similarity, 162
  trait evolution, 79–82, 84–86, 409
  See also plesiomorphy
ancestral gene, 151–152, 155
ancestral lineage
  branching of, 40, 44–45, 53, 372
  in paraphyletic groups, 125
  in polytomies, 58–59
  unrooted trees, 63
ancestral node
  in clade definitions, 44
  traits of, 335–337
ancestral range, 353
ancestral sequence, 88, 106, 197, 224
ancestral state
  continuous traits, 337
  in Hennigian inference, 179, 416
  in tests of adaptation, 306
  in trait evolution, 85, 163
  plesiomorphy, 116
  reconstruction, 308, 312, 317–319, 340
ancestry walk, 2
animal phylogeny, 60–61, 70–71, 390, 397
animal pollination, 323–325, 423
Antarctica, 19–20, 353–355
antorbital fenestra, 116–117
apes
  genomes of, 295
  humans as, 30, 32, 405
  shoulder evolution, 310
aphids, 42–43
apomorphy
  association with clades, 85, 179
  definition, 85, **439**
apomorphy-based clade definition, 131
Aquilegia, 377
Archaea, 161
Archie, James, 21–22
archosaurs, 61–63
arthropods
  example of a species rich clade, 371–372
  phylogeny, 70–72
artificial classification, 119
asexual clones, 80
asexual organisms, 42–43, 140
asexual reproduction
  effect on diversification, 388, 423
  features of, 42–43
asymmetric model, 320–321
asymmetric rate matrix, 320–321
auditory bulla, 175–176
Australasia, 20, 353–355
autapomorphy, 185, 416, **439**
autocorrelated relaxed clock. *See* correlated relaxed clock
Aves. *See* birds
backcrossing, 159
bacteria, 3, 41, 161, 386
baculum, 176
balanced tree, 373
baobab (*Adansonia*), 56, 100

base composition
inequality of, 225–227, 238
stationarity of, 218
See also base frequency
base frequency
Bayesian inference, 250
maximum likelihood, 242
model of, 225–229, 418
model selection, 246–247
base pairing, 7, 86, 230
base substitutions, 196–199
bats
evolution of echolocation, 97, 99
homology of wing, 3, 15, 17
relationships to other mammals, 278–279
SARS virus, 6
species tree example, 294
Bayes' factor analysis, 285
Bayes' theorem, 248, 250, 261
Bayesian comparative methods, 312, 325–327
See also Stochastic Mapping
Bayesian concordance analysis, 295–296
Bayesian MCMC. *See* Markov chain Monte-Carlo
Bayesian phylogenetics
comparison to other methods, 195, 207, 217–218, 238, 249, 258, 316, 419
hypothesis testing, 284–285, 303, 420
molecular dating, 362, 365–369
tree estimation, 247–259, 262
Bayesian posterior probability. *See* posterior probability
Bayesian species tree methods, 292–294
Bayesian statistics, 247–250, 257, 417, **439**
BCA. *See* Bayesian Concordance Analysis
bears
biogeography, 19–20
molecular dating example, 386–387
species concepts, 165
See also carnivorans, worked example
BEAST, 366
beetles, 282–283
behavior
divergence of, 39, 160, 163
for phylogenetic inference, 395
heritability of, 3, 79, 83
male sacrifice, 313–315
nest building, 116
Bessey, Charles E., 23–24
biased coin
Bayesian example, 248–249, 253
likelihood example, 239–240
bilateral flower symmetry. *See* zygomorphy
Bilateria, 60–61
binary character, 316–317, 321, **440**
binary tree
definition, 58, 63, **440**
tree balance, 372
binary-state speciation and extinction (BiSSE), 379–381, 388, **440**
binomial, 128, **440**
biochemical traits
divergence of, 160
for phylogenetics, 79, 83, 115, 395–396
biogeography
as evidence for common ancestry, 16, 19–20, 405
for phylogenetic inference, 396
heritability of, 20, 79
studying with phylogeny, 351–355, 369–370, 381
biological classification. *See* taxonomy
biological species concept, 164, 171, 413, **440**
bipartition, 63, **440**
See also split
birds
apomorphies of, 85, 116
as dinosaurs, 4, 138, 411
broader relationships, 44, 132
diversification rate, 375
evolution of flight, 6–7
evolutionary advancement, 10–11
homology of wings, 90
malaria vector, 385
ratites, 353–355
relatedness to other reptiles, 27, 33, 45, 85, 115–117, 406
SARS virus, 6
taxonomic history, 120–125
See also feathers; wings
BiSSE. *See* binary-state speciation and extinction
body size
as continuous trait, 327–328, 330
correlations to other traits, 328, 338
bootstrap
non-parametric, 273–281, 297, 403, 420, **450**
parametric, 281–283, 297, 365, **451**
bootstrap consensus tree (bootstrap tree)
example of, 60, 277–280, 354, 358
interpretation, 275–277, 280, **440**
bootstrap replicate, 273–276, 300–301, 420, **440**
bootstrap score (bootstrap percentage)
meaning, 60–61, 74, 265, 275, 277, 284, 301, **440**
on trees, 60, 277–280, 282–283, 354, 358, 407
botanical code, 130
branch length
as evolutionary distance, 236, 238, 241–242, 245
ignored by parsimony, 205, 316
in Bayesian inference, 250–252
in comparative methods, 305, 312, 316–317, 322, 335, 337, 346, 421–422
in distance methods, 236–238, 418, 448
in maximum likelihood analysis, 240–245, 417
in parametric bootstrap, 281
in phyloecology. *See* cumulative evolutionary age
in species tree methods, 294
in tree-to-tree distances, 65
relation to time, 5, 54–55, 362–364, 369
representing amount of change, 56, 74, 115
under pure birth model, 374, 378

branch-and-bound, 189–190, 193, 200, 436
branch-based clade definition, 132–133, 136, **440**
branch-to-branch rate heterogeneity. *See* rate heterogeneity
Bremer support. *See* decay index
Bremer, Kåre, 272
Brown, Robert, 330
Brownian motion, 330–339, 343, 346, 421, **440**
bulletin board, analogy for molecular dating, 366
burn-in, 254–255, 367, **440–441**
butterflies, wings of, 121
cacti, geographic distribution, 20, 31, 405
calibrated tree, 5, 362–363
calibration, 369–370, **441**
camel, gene duplication example, 154–157
caniforms, 201
canine teeth, 174
cards
  metaphor for bootstrap, 273–275
  metaphor for sequence evolution, 219–227, 234
carnassial teeth, 174
carnivorans
  biogeography example, 369–370
  independent contrasts example, 338–339
  tapeworm host, 356–357
  worked example
    assessing data set conflict, 285, 288, 291–292
    background, 173–174
    Bayesian inference, 255–258
    bootstrapping, 280
    distance methods, 232–238
    generating a matrix, 173–178
    Hennigian inference, 179–182
    maximum likelihood, 244–245
    parsimony, 192–194, 200–202, 204–205
    phylogenetic signal, 270–271
carnivory, 347, 422
cartoon, 28–29
cats
  biogeography example, 351–352
  in mammal phylogeny, 278–279
  species tree, 304
  tapeworm host, 357
  See also carnivorans, worked example
CEA. *See* cumulative evolutionary age
cells, trees of, 42
cetaceans, 11, 347, 422
  See also whales
character conflict, 266
character encoding, 395, 399–403, **441**
character history, 97, 325–326, 423
character polarity, 84–86, 100, 175, 179, **441**
character scoring, 181, 399, **441**
character state change, 79–84, 89, 93–94, 100, 182–184, 202–205, 378, 409
character state dependent diversification, 376, 379, **441**
character state matrix, 175–177, 237, 240, 250, 418, **441**
  See also data matrix
character state weighting, 202–204, 214, 392, 416, **441**
character state, as contrasted to character, 80, 88, 175–177, 217, **441**
character weighting, 202–204, 207, 416
character, as contrasted to character state, 80, 88, 175–177, 217, **441**
characters for phylogenetic inference, 175–176, 389–397
chi-square distribution, 247, 364
chi-square test, 312–313
chimpanzees
  close relationship to humans, 3–4, 110
  example of relatedness, 137, 411
  gene tree example, 150–151, 295
  hunting, 310
  implied ancestor to humans, 28–29
  relationships to other primates, 73, 133, 295, 309–310
  tool use, 309
chromosomal inversions, 393–394
chromosomes
  gene order, 393
  in sexual reproduction, 140–146
chronogram
  absolute, 54–56, 362–363, 369, 386–387, **439**
  definition of, **441**
  in studies of diversification, 374, 380
  relative, 362–369, **454**
CI. *See* consistency index
circle tree, 48, **441**
civet
  SARS virus, 6
  See also carnivorans, worked example
clade
  age of, 131, 164, 375
  association with synapomorphies, 85, 116
  basis of taxonomy, 122–124, 126–133
  content of cladograms, 54
  definition of, 44–45, 121, **441**
  exclusivity of, 122–123
  for comparing trees, 46–48, 51–52
  in consensus trees, 59
  merging of, 52, 164, 448
  pruning of, 49–52
  ranking, 128–131, 137
  tree-to-tree distance, 66–67
  unrooted trees, 61–62
clade credibility, 60, 257–258, 273, 386, **441**
clade names, 127, 412
  See also clade definition
clade posterior probability. *See* clade credibility
clade size, 349, 370–375
clade support, 60–61, 265, 271–273, 281, 297
cladists, 27
cladogenesis, 38, **441**
cladogram, 54, 56, 115, **441**
class, group of organisms, 12
class, taxonomic rank, 27, 121–123, 128, 130, 135

classification
based on evolutionary relatedness, 25–28, 107, 117–129, 411
evidence for evolution, 16, 20–21
interchangeable with a tree, 7, 126–128, 137
climate change, 2, 352
clock model. *See* molecular clock model
clock. *See* molecular clock
clustering, phylogenetic. *See* phylogenetic clustering
Cnidaria, 60–61, 70–71
co-occurrence, of species, 359–361, 423
coalescence
in branching population trees, 147–150
in species trees, 292–294, 304, 420–421
of gene trees, 142–144, 166, **442**
coalescence model, 294
coalescence theory, 293, **442**
coalescence units, 294, **442**
coalescence, deep. *See* deep coalescence
code, genetic, 15–16, 84
code, nomenclatural, 130–133, 412, **442**
coding genes, 198, 203, 286, 390–391
codominant tree, 160, **442**
codon models, 200, 203, 230, 286
coelacanths, 29, 114, 123–124, 375
coevolution
of different species, 355, 357, 381
of interacting bases, 390
coin toss
Bayesian example, 248–249, 253
likelihood example, 238–240
Markov model, 218
cold chain, 255
Coleoptera, 282–285
collapsing branches, 280
columbines, 377–378
combined alignment and tree building, 197
combined analysis, 286–288, 291, **442**
common ancestor
depiction in trees, 44, 126–127, 131
general concept, 3–5, 406
geographic distribution of, 352
historical significance, 11–14, 19, 406
homology of traits, 90–91, 409
in phylogenetic systematics, 25, 406
non-monophyletic groups, 124–125
recency of, 27, 108–115, 117, 134, 259, 386–387, 406
traits of, 28–29, 90–91, 136, 186, 332, 334–335, 407, 411
common ancestral allele, 142, 152–153
common ancestry
acceptance of, 23, 41
basis of homology, 91
basis of relatedness, 4, 107–110, 117, 185, 268
definition, 12–15, **442**
evidence for, 15–23, 32, 201, 286, 350
explanation of similarity, 334
importance of, 3–4, 14, 31, 77, 271
in history of evolutionary theory, 9–23, 31
in phylogenetic inference, 221, 397
in phylogenetic systematics, 27–28, 107, 125, 406
predicts agreement among trees, 286
common descendant, 147
community composition, 359–362, 384, 423
comparative method, the
continuous traits, 337–339
discrete traits, 311–315
general concepts, 311, 422, **442**
comparing data sets, 285–288, 297
compatibility
of tree topologies, 50–52, 73, 284–285, 287–288, **442**
of trees with data sets, 173, 210, 442
plant sexual system. *See* self-compatibility
reproductive, 164
competitive exclusion, 360–361
complementary base pairing, 86, 397
complete lineage sorting, 150
See also incomplete lineage sorting
con artist, analogy to bootstrap, 273
concatenated data, 286, **442**
concentrated changes test
model-based analog, 327
parsimony, 313–316, 340, 347, 422
concepts of species. *See* species concepts
concordance factor, 150, 168, 295–296, **442**
confidence interval, 284, 322, 336, 360, 365, 368
confidence threshold, 240
confidence, in clades, 60, 247, 265
conflict (disagreement)
among characters, 200, 214, 266–267, 415, **442**
among gene trees, 150, 160, 292, 420
between data sets, 286–291, 357
See also discordance
consensus tree
Bayesian majority rule, 257–258, **439**
bootstrap majority rule, 275, 277, 280, **440**
strict, 59, 193, 287, **442**
consistency, 93–95, 100, **442**
consistency index (CI), 21–22, 93–95, 97–98, 100, 105, 182, 193, 269, **442–443**
consistent characters, 94, 181, 200, 202, 415, **443**
conspecific, 164
constant rate of diversification, 374
constrained search, 300, **443**
constraint tree, 282–283, **443**
continental drift, 352, 355
contingency table, 312, 328
continuous traits
evolution of, 305, 327–340, **443**

test for correlated evolution, 305, 311, 328–329, 336–340, 347
continuous variation, scoring as discrete, 402
continuous-time Markov models, 218, 230
contrasts. *See* independent contrasts
convergence, 92–93
convergent evolution, 91, 125, **443**
converting classifications into trees, 126–128
converting trees among formats, 48, 67
cophylogenetics, 355–357, 381, **443**
correcting pairwise distances, 234, 236
correlated evolution
continuous traits, 305, 327–329, 336–340, **443**
discrete traits, 305, 311–313, 323–327, 340, 343, 346–347, 421–422, **443**
correlated relaxed clock, 367, **443**
cospeciation, 356–358, **443**
cost
gap, 198, **445**
in weighted parsimony, 203–205, 246, 315, **443**
cotton, 56
coupled chains, 255
cow
gene duplication example, 154–157
relationships of, 135, 278–279, 347
SARS virus, 6
creation, independent or special, 15–16, 21
credibility intervals
on concordance factors, 150, 295–296
on nodal depth, 365, 368
creodont, 175
See also carnivorans, worked example
crocodiles
archosaur phylogeny, 62–64
broader relationships, 111–114, 123–124, 134, 340
diversification rate, 375
relationship to birds, 27, 33, 45, 85, 115–117, 406
crown node (crown group), 131, **443**
cumulative evolutionary age (CEA), 360–361
Darwin, Charles, 9–15, 18–23, 33, 113, 119–121, 349, 406
Darwin, Erasmus, 11–12
Darwin's finches, 344
data matrix
conversion to distance matrix, 232, 237
features of, 175–176, 196, **443**
frequency data, 396
molecular, 196, 209
morphological, 173–182, 193, 399–403
partitions in, 286, 291
data set conflict. *See* conflict, between data sets
dated fossil, 54–55, 287
dated nodes, 362, 369
daughter lineages, 44–45
daughter populations, 39, 147
daughter sequences, 86
decay index, 265, 271–273, 284, 300, **444**
deep coalescence, 148–150, 292–293, 304, 420–421, **444**
See also minimizing deep coalescences
degree of relatedness, 4, 25, 27, 108–110, 114, 117
degrees of freedom, 247, 364
deletions, 88–89, 177, 196–197, 393, **444**
dependent model, 323–324
depth
nodal, *See* nodal depth
of trait mapping, 84
derived allele, 77–79, 84, 409
derived character state (derived state)
as adaptation, 306–308
association with clades, 84–85, 162, 179–180, 186, 416
autapomorphy, 185
character polarity, 84–85, 88
consistency, 93
effect on diversification, 376
evolution of, 79–80, 83, 409
See also synapomorphy
descendant lineage
after gene duplication, 152, 159
in lineage fusion, 159, 447
in lineage splitting, 39, 82, 447
in tree diagrams, 45, 58, 63, 111
trait evolution, 80, 82, 91, 332
tree balance, 372
vicariance, 362
descent from common ancestry. *See* common ancestry
descent with modification, 77, 80
development thinking, 28
See also ladder thinking
developmental correlation, 311
dewclaw
homology example, 106, 409
in carnivorans, 176, 179
Dewey Decimal system, 118–119
diagnostic traits, 125, 165, **444**
diagonal format, 43, 48, 58, 73, **444**
diagonal-down format, 48
diagonal-up format, 48
diagonals, in rate matrix, 226–227
dichotomous, 58, **444**
dinosaur, 4, 6, 62, 124, 138, 411
diploid, 140–141, 151, 166, 412
Diptera, 282–283
directed graph, 43, **444**
direction of trait evolution. *See* polarity
directional trend, in trait evolution, 320–322, 325, 379, 421, **457**
discordance
between gene trees, 143–145, 150–151, 161, 295–297, 421, **445**
between population and gene trees, 149–151, 171, 292–295, 420
discordant clades, 66
discrete approximation, 229
discrete character. *See* discrete trait
discrete traits
ancestral state reconstruction, 317–319, 340
comparison to continuous traits, 327–329
model of evolution of, 316–317, 340
scoring of, 395, 399–402
testing for correlation, 323–325, 340, 347
dispersal (migration)
analogy to host-switching, 356

dispersal (migration), *continued*
concept of, 39, 79, **444**
rate of, 19, 79, 349–350
reconstructing history of, 349–352, 355
role in lineage splitting, 39, 362, 376
to Hawaii, 350–351
to Madagascar, 369–370
wind, BiSSE example, 388
distance data, 237, 397
distance matrix
analysis of, 234–237
construction of, 232–234
frequency data, 397
loss of information, 237
distance methods
for phylogenetic inference, 231–238, 259
pros and cons, 217, 397, 418
distance optimality methods, 237–238
distance, tree-to-tree, 65–67, **457**
divergence
of gene duplicates, 5, 43, 151, 155, 157
of lineages, 2, 39–40, 142, 161, 266
of trait values, 332–334
rapid, 420
diversification
causes of, 371, 375–381
character-state dependent, 378–381, 388, 423, **441**
effect on comparative methods, 322
effect on ranking, 131
rate, 370–375, 381, **444**
DNA sequences
aligning, 195, 200–202
definition of gene, 146
distance analysis, 231–238
evolution of, 86–89, 218–231
example of consistency, 93–94
homology of, 90
likelihood analysis, 239–245
parsimony analysis, 195–201, 204, 212, 214
partitions, 285–286
related to gene trees, 140, 145, 157–161
use for phylogenetic inference, 175, 177–178, 389–391
DNA-DNA hybridization, 238, 397
dogs
homology example, 106, 409
See also carnivorans, worked example
dolphins, 80, 347, 422
dominant history, 158, 161, **444**
dominant tree, 158–160
dorsal scapula, 308–310
downhill proposals, 251
drift, of trait values, 330, 349
Drosophila, 74, 282–283, 394, 407–408
duplicate genes, 5, 151, 157
duplication. *See* gene duplication
duration, of branches, 45, 54–55, 56, 218, 241, 315, 367
ears, 176, 179, 204
echolocation, 97, 99
ecology
for phylogenetic inference, 396
importance of tree thinking in, 2, 339
phylogenetic approaches to, 359–362
edge, 43–44, **444**
See also branch
effective mutation rate, 226–227
effective population size, 150
elephant, 52, 278–279
embryological data, 23, 25
encoding, of characters, 395, 399–403, **441**
epidemiology, 6
epithet, specific, 128, **455**
equal rates model. *See* symmetric model
equally parsimonious reconstructions, 97, 99, 183, 316, 352
equally parsimonious trees, 286–287
equally weighted parsimony
ancestral state reconstruction, 309–310, 352, 385, **444**
phylogenetic inference, 202–205, 215, 268, 416, **444**
equivalent trees, 45–48
equivocal reconstruction, 315, 351
error
in posterior probability, 284
in reading trees, 111–112
in tree reconstruction, 23, 181, 265, 416
sampling, 281
estimate
of age, 54
of ancestral state, 327
of branch length, 241
of the true tree, 186, 240, 265
of value, under Brownian motion, 332
evidence for evolution/common ancestry, 9, 14–33, 201, 268, 286
evidence of selection, 311–312, 340
evolution
as an ongoing phenomenon, 41
definition of, 78
non-progressive nature, 28–30, 113
evolutionary advancement, problems with, 9, 23, 25, 46
evolutionary age, 131
evolutionary chronicle. *See* historical chronicle
evolutionary correlation, 312–315, 328, 340, **444**
evolutionary distance, 221–225, 230–237, 241, 260, 363, 367, **444**
evolutionary history
organisms as summations over, 83
using trees to study, 2–4, 305
evolutionary kinship, 21, 25, 50, 163
evolutionary lineage. *See* lineage
evolutionary narrative. *See* historical narrative
evolutionary systematists, 27, 32, 405
evolutionary timescale, 37–38, 54
evolutionary trend, 320, 379
exaptation, 310–311, 341, 421, **444**
exclusive groups, 122–123, 162–164
exclusivity, 122–123, 162–164, **444**
exhaustive taxonomy, 126
exponential distribution
for uncorrelated relaxed clock, 366
of waiting times, 219, 372
external branch, 44, 440, **445**

external specifier, 132–133, 136, **445**
extinction
effect of key innovation, 376
in biogeography, 355
of allelic variants, 230
of gene copies, 139, 151, 156–158
extinction rate, 371–372, 375, 388
extrinsic isolation, 39
eyeball alignment, 199
F81, model of evolution, 225–229, 242, 246–247, 262, 418
fairy, card changing, 219–221, 224–227, 234
families, of Carnivora, 174, 351, 357
family, taxonomic rank, 128–130, 136
family tree, 35, 45, 107
father, 109, 141–143
feathers, 6, 85, 117–118, 120–123
feliforms, 200
Felsenstein, Joseph, 1, 273, 311
figs, 356–358
fins, 28, 80
fish
accelerated diversification, 375
"primitive" status, 11, 28–29
sympatric speciation, 39
taxonomic history, 117, 121–124
swordtail, 319–320, 326
fissipeds, 174
Fitch parsimony, 202, 204–205
See also equally-weighted parsimony
Fitch, Walter, 202
fixation
in relation to adaptation, 306
of derived alleles, 78–80, 84, 158
of derived traits, 131, 376, 409
flamingos, 306
flat prior probability, 250, 285, **445**
flat-weighted parsimony. *See* equally-weighted parsimony
flight, evolution in birds, 6–7, 116
flower color, 7
flower showiness, 323–325
flower symmetry, 321–322, 388
flowering plants (Angiospermae), 23–24, 137, 205, 303, 341, 410
football, metaphor for trait evolution, 330–336
forensics, 6, 202
fossil calibration, 369–370
fossils
as tips, 35, 41, 49
evidence for evolution, 2, 16–19, 32–33, 405
morphological data, 403
uncertain traits, 176, 399
use in dating, 54–55, 362, 369
vertebrates, 123–124
fragmentation of geographic ranges, 92, 352, 355
free parameters, 229, 246–247, 250–251, 364
frequency data, 396–397
frogs
Bayesian inference example, 261, 417
relationships of, 102, 110, 124
fully resolved, 58, 65, 187, **445**
See also binary tree
function
for which a trait evolved, 3, 306
in relation to adaptation, 3, 306–311, 421
in relation to evidence for evolution, 15–16
of a gene, 146, 151, 154, 230
functional constraint, 286, 390
functional gene, 146
functional morphology, 339
fur
comparative methods example, 311–313, 343
mammal trait, 117–120
See also hair
fur color, 312–313, 399
fynbos, 360–361
gain-loss bias, 320–321
gains, of traits, 97–99, 320–321, 327, 392
gall wasps, 318
gamma distribution, for rate heterogeneity, 229–230, 256
gap penalty (gap cost), 198, **445**
gaps, 177–178, 196–200, 218, 415, **445**
gene, functional, 146
gene, recombinational, 146, 151, 161, 292, **453**
gene duplication, 5, 139, 151–157, 166, 169–171
gene family, 5, 151, 154–157, **445**
gene flow
between populations, 38–39, 42, 78, 159, 352, **445**
limits to, 356, 376
gene genealogy, 141, 144
gene history, 139, 142–145, 148–151
gene lineage, 142–143, 147–149, 153, 156, 158
gene loss, 157
gene order, for phylogenetic inference, 393
gene trees
in general, 89, **445**
in population trees, 139–171
in species tree methods, 292–294, 304, 420
indel events on, 197, 199
of gene families, 5
gene-to-gene discordance. *See* genealogical discordance
genealogical discordance
causes of, 143–145, 149–151, 171
definition of, 139, **445**
importance of, 161,
methods for studying, 292–297, 420–421
genera. *See* genus
general time reversible model, 228–229, 246, 256, 418
generalized least squares, 339
generalized parsimony (weighted parsimony), 203–205, 207, 215, 246, 315, 390, **445**
genes, as tips, 35, 44
genetic code, 15–16, 84
genetic drift, 78, **445**
genetic locus, 140–141
genome duplication, 151
genus, taxonomic rank, 128, 165, **445**
geographic distribution
analogy to host-parasite relations, 356
cause of lineage splitting, 39, 82, 147
community composition, 359, 362
evidence for common ancestry, 16, 19–20
heritability of, 3, 79

geographic distribution, *continued*
reconstructing past changes in, 349–355, 369–370
showing on tree, 83
trait for tree inference, 396
geological events
cause of lineage splitting, 39
explaining species distributions, 353–355
gibbons, 73, 132–133, 278–270, 309–310
global optimum, 191–192
globally optimal tree, 272
Gondwana, 353–355
gorillas
relationships to humans and chimpanzees, 137, 150–151, 295, 411
relationships to other primates, 73, 132–133, 309–310
tool use, 309
grade. *See* paraphyletic group
grandparent, 41, 108–109, 114
Great Chain of Being, 2, 10
See also *Scala Naturae*
GTR. *See* general time reversible model
Haeckel, Ernst, 25–26
hair
example of adaptive hypotheses, 307
example of character, 196
example of developmental correlation, 311
example of tip correlation, 312
mammal trait, 80, 122
hallux, 176
haploid, 151
haplotype. *See* allele.
hard polytomy, 58–59, 266, **446**
Hawaii, 350–351
heated chain, 255
hemoglobin
branch length example, 56
gene duplication example, 169
Hennig, Willi
in phylogenetic systematics, 25, 27, 32
terminology coined by, 85–86
Hennigian argumentation, 178
Hennigian inference
assumptions of, 179, 416
comparison to parsimony, 181, 416
historical significance, 173
limited use of, 217
method and application, 178–181, 207, 208, 215
problems of, 180–181
structural molecular characters, 393
herbarium specimens, 166
heritability
needed for tree inference, 175, 389, 395
of geography, 349
of niche, 360
of selective regimes, 308
of traits, 79, 83, 90, 395, **446**
Herron, Jon C., 80–81
heuristic search, 190–192, 212, 237, 252, 416, **446**
hidden changes, in substitution model, 219, 224
hierarchical classifications
as evidence for evolution, 20–21, 119–120, 128
naturalness of, 118–120
hierarchical likelihood ratio test. *See* likelihood ratio test
hierarchical Venn diagram, 128
hill-climbing algorithm, 191
See also heuristic search
Hippopotamus, 135, 278–279, 347, 422
historical biogeography. *See* biogeography
historical chronicle, 53, 305–306, 347, 422, **446**
historical narrative, 305–307, 315, 345, 347, 375, 422, **446**
historical sciences, 173, 202
historical signal, test for, 266
history of systematics, 23–28, 120–124
HIV, 6
HKY, 228–229, 244, 246, 262, 268, 418
home range, 338–339
Homeothermia, 124–125
homeothermy, 124–125
Hominoidea, 132–133
homologous gene, 141, 152, **446**
homologous positions, 89, 178, 195–197
homologous trait, 3, 17, 106, 409, **446**
homologs, 152, 197, **446**
homology
contrast to homoplasy, 93
errors in, 416
of gene copies, 153
of nucleotide positions, 86, 90
of phenotypic traits, 90–92, 100, 106, **446**
homoplasy
cause of character conflict, 266
concept of, 93, 100, **446**
effect on weighting, 202–204, 416
in branch-and-bound searches, 189
in Hennigian inference, 179, 181
in parsimony, 181, 186, 215, 416–417
of limb evolution, 121
on long branches, 207
relation to consistency index, 94, 97
signs of treelike structure, 266
structural molecular characters, 393
homoplasy index, 94
homozygotes, 77
horizontal gene transfer. *See* lateral gene transfer
horses
gene duplication example, 154–157
limb bone homology, 15, 17
relationships of, 134–135, 278–279, 347
host-parasite coevolution, 356–358, 385–386
host-switching, 356–358, 385–386
Huelsenbeck, John, 282–283, 346
humans
adaptation examples, 307–311, 340–341
example of common ancestry, 3–4, 12, 14, 41
example of relatedness, 110–114, 410
genealogical discordance, 150–151, 295
homology examples, 15, 17, 90, 106, 409
limb bone homology, 15, 17
meat eating, 356

relationships to chimpanzee/gorilla, 28–30, 137, 150–151, 295, 411
relationships to other animals, 132–134, 278–279
relationships to other primates, 73, 133, 309–310
species name, 128, 165
supposed advancement of, 9–10, 25, 28–30, 405
tapeworm hosts, 356–357
utility of phylogenies to, 5–6
hummingbirds, 7, 20
hunting with projectiles, 84, 309–310
Huxley, Thomas Henry, 108–109
hybrid speciation, 139, 159–160, **446**
See also lineage fusion
hybridization
DNA-DNA, 238, 397
effect on gene trees, 158–160, 167
effect on population history, 158–160
in tomatoes, 295–296
limits on, 40, 160–161
Linnaeus' views on, 128
hyena
tapeworm host, 357
See also carnivorans, worked example
immunological distance, 397
incisors, 174, 176
incompatibility. *See* self-incompatibility
incompatible tree topology, 52
incomplete lineage sorting
cause of genealogical discordance, 139, 149–151, 166, 286, 292–296, 413, 421, **446**
molecular presence/absence data, 392
incongruence length difference test. *See* partition homogeneity test
inconspicuous flowers, 323–324
indels
calculating pairwise distances, 232
in sequence alignment, 89, 197–200, **446**
models of, 230
treated as missing data, 214, 218
indented list, 126–128, 137
independent contrasts, 337–339, 344, 422
independent model, of trait evolution, 323–324, 364
independent origins
in Lamarck, 10
of molecular structural traits, 394
independently created, 15–16, 21
India, 353, 382–383
inference
Bayesian, 247–259, 284–285
Hennigian, 173, 178–181, 207, 208, 215
in science, 7, 25, 95, 173, 202, 265, **446**
of trait evolution, 95–99, 305–340
phylogenetic, 173–263
Inferring Phylogenies (Felsenstein), 1
information storage, 5, 83–84
informative trait, 185, 266, 209, 266, 415
See also parsimony-informative
ingroup, 64, 175–176, 179, 187, **446**
inheritance of acquired characters, 406
input tree, 59, **447**
inquiline wasps, 318
insects
evolution, 95–98
in Lamarck, 11
parametric bootstrap example, 282
relationships to other animals, 71
wing homology, 3, 90, 95–98
insertions, 88–89, 177, 196–197, 393, **447**
instantaneous rate matrix, 221, 226, 228
instantaneous rates, 220–221, 226–228, **447**
interbreeding, 146, 159, 164, 413
internal branch
annotated with support, 61, 74, 275, 277
collapsing, 59
definition, 44, **447**
in unrooted trees, 61, 63, 194
rotation around, 46
internal node
age of, 54
ancestor of clade, 47, 127
assessing relatedness, 112
state assignment, 243, 325, 335
internal specifier, 132–133, 136, **447**
internode, 44, 54, 294, **447**
See also internal branch
intrinsic isolation, 39
introgression, 139, 157–161, 167, 286, 292, 296, 413, **447**
introns, 203, 286, 393
invasion
biogeographic, 350–352
of land by animals, 28, 124
of land by plants, 137
of water by whales, 422
inversions, 196, 393–394
irreversibility of fixation, 78
isozymes, 391
jaguar, 304, 351
JC model
comparison to other models, 225–227, 229–230, 246–247, 262, 418
in distance analysis, 234–236, 238
in likelihood analysis, 240–243
model description, 224–225, 259
similarity to models of discrete trait evolution, 316
Jukes-Cantor model. *See* JC model
kangaroos, 20, 114, 132, 278–279
key innovation, 375–379, 388, **447**
King, Larry, 30
kinship, 21, 25, 27, 50, 108, 120, 163, 282, 406
See also relatedness
ladder of life, 2, 10, 12, 23, 25, 30–31, 113, 406
ladder thinking, 9, 28–30, 32–33, 113, 406, **447**
Lagrange, 355
Lamarck, Jean Baptiste de, 9–12, 33, 406
lampreys, 29, 123
land crabs, biogeography example, 382
land plants, 137, 394, 410

last common ancestor
clade age, 131
distance from the present, 41, 84, 110
features of, 28, 90–91, 352, 409
in gene trees, 142, 152–153
in relatedness, 109–114, 411
on a tree, 44, 126–127
lateral gene transfer (LGT), 139, 157–161, 167, 413, **447**
leaf length, 330, 346, 400–403
leaf shape, 400–402
least-squares method, 237, 260
leaves, 21
See also tips
lemurs, 278–279, 369–370
length. *See* tree length or branch length
leopard, 304, 351–352, 421
less parsimonious, 97, 186, 189–190
LGT. *See* lateral gene transfer
likelihood
definition, 239, 248, 261, 417, **447**
for ancestral state reconstruction, 316–322, 340, 342
in Bayesian inference, 248–254, 262
in maximum likelihood tree inference, 239–244, 268, 272–273, 289
in model selection, 246–247, 254, 262, 418
in molecular dating, 363–365, 367
in Pagel test, 323, 325
likelihood function, 242
likelihood ratio, 239, 272–273, **447**
likelihood ratio test, 247, 254, 320, 323, 364, **447**
likelihood score, 261, 272, 289, 320–322, 417
limbs
evidence for evolution, 15, 17
in vertebrate taxonomy, 117–118, 120–125, 136
secondary loss of, 80, 347
lineage birth rate, 371, **447**
lineage branching, 45, 82, 334, 409, 417
lineage death rate, 371, **447**
lineage diversification rate, 350, 371, 377, 381, 423
lineage fusion, 157, 159–161, 292, **448**
lineage sorting, 143, **448**
See also, complete lineage sorting; incomplete lineage sorting
lineage splitting
causes of, 39, 82, 147, 355–356, 377
definition, 39, **448**
in gene trees, 153
polytomous, 58–59
rate of, 79, 370–372, 376–377
lineal ancestors, 40–41
lines of descent, 37, 109, 139, 141
linkage, 145
Linnaeus, Carl, 128
lions, 304, 351–352, 421
listing clades, to describe trees, 47, 63
living organisms, as ancestors, 24, 28–29
lizards
diversification rate, 375
example of trait evolution, 80–83
example of PGLS, 339–340
malarial vector, 385
relationships to crocodiles/birds, 27, 33, 45, 85, 115–117, 124, 406
lobe-finned fish, 29
See also coelacanth; lungfish
Lobeliaceae, 350–351
local optima, 191–192
locus (loci), 77, 79, 140–144, 396, 409, **448**
log-likelihood
definition, 241, 243–244, 261–262, 268, 272, 364, **448**
ratio, 239
See also likelihood ratio
log-normal distribution, 366, 368
long branch attraction, 206–207, 282, **448**
long-distance dispersal, 349–350, 356, 376
losses, of traits, 78, 96–99, 320–323, 347, 379, 388, 392, 422
lungfish
diversification rate, 375
relationships to tetrapods, 102, 110–114, 123–124
lungs, 114, 117–118, 120–123
Lyell, Charles, 12
MacClade, 315
Madagascar, 100, 353, 369–370
majority rule consensus tree, 257–258, **448**
malaria, 385
male genital mutilation, 313–315
male sacrifice, 313–315
Mammalia, 27, 121, 123, 132
mammals
"advancement" of, 10, 113
carnvorans, 174–175
evolution of homeothermy, 124–125
example of independent contrasts, 338–339
example of molecular dating, 369–370
internal phylogeny, 51, 99, 132, 278–280, 368
malaria hosts, 27
monophyly of, 27
possession of hair/fur, 80, 84, 307, 311
rate of diversification, 375
relationships to other amniotes, 45, 102, 115–117, 120–124
SARS virus, 6
manual alignment, 199
mapping
character state changes on tree, 83–84, 120, 162, 183–184, 207, 308–310, 313–315, **448**
geographic distributions on tree, 382, 396
restriction sites, 393
selective regimes on tree, 313
stochastic. *See* stochastic mapping
marking trait changes on a tree, 82
Markov chain Monte Carlo (MCMC), 250–257, 261–263, 284–285, 312, 327, 364–368, 417–420, **448**
Markov chain Monte Carlo generation, 252, **448**
Markov chain Monte Carlo proposal, 251–253, 261
Markov model, 218, 230, **448**
Markov process, 218, 329
marsupials, 280, 355, 368
maternal allele, 109, 141, 143, 145

mathematical models (probabilistic models), 207, 217–218, 240, 316, 323, 392
maximum expected pairwise distance, 234
maximum likelihood
bootstrapping, 275, 281
comparison to Bayesian inference, 250, 254, 263, 284, 419
comparison to parsimony, 195, 202, 207
for phylogenetic inference, 238–245, 259, 356
for reconstructing trait evolution, 316–325, 329, 340, 342, 378
for RFLP data, 392
for species tree estimation, 292–294
general principles, 239–240, **448**
model selection, 246
models for, 217–231
molecular dating, 362–365
partition homogeneity, 288
phylogenetic signal, 268
maximum parsimony
bootstrapping, 275
need for statistics, 297
phylogenetic inference, 173, 181–187, 195, 200–207
reconstructing trait evolution, 95–100, 313
See also parsimony
MCMC. *See* Markov chain Monte Carlo
meiosis, 143
merging clades, 49, 51–52, **448**
Mesquite, 46
mice
as "lower" mammals, 113
common ancestry with humans, 12, 14
genealogical discordance example, 145–146
relatedness example, 108, 110–114
relationships, 278–279
microsatellites, 392
migration, 19, 350–352, 355, 369–370
See also dispersal
milk, 117–118, 120–123
Milligan, Brook, 323
minimal age, 369
minimizing deep coalescences (MDC), 292–293, 304, **448**
minimum evolution, phylogenetic inference method, 237–238, 260, 288, 418, **449**
minor history. *See* minor tree
minor tree, 158–159, 413, **449**
missing data, 200, 214, 218
See also uncertainty
mitochondrial DNA (mtDNA), 303, 390–391
mixing, 254–255, **449**
ML. *See* maximum likelihood
moas, 355
mobile genetic elements, 196–197
model of coin tossing, 218, 239
model parameters, 281
model selection (model choice), 246–247, 254, 259
model-based methods
for biogeography, 355
for phylogenetic inference, 217–259, 218, 363
for species trees analysis, 293–294, 296
for trait evolution, 305, 316–327, 347, 422
general features, 218, 246, **449**
models of evolution
comparing partitions, 286
for phylogenetic inference, 224, 230, 242–244, 245–259, 363
for trait evolution, 317, 324
molecular clock, 64, 115, 230, 246, 363–366, **449**
molecular data
analysis of DNA sequences, 195, 200, 205, 209–210, 236–238, 291
carnivoran example, 200–201, 205, 236–238, 244–245, 255–258, 271, 280, 288
comparison to morphology, 21, 88–89, 201, 257–258, 271, 280, 288, 291–292, 399
nonsequence molecular data, 391–397
molecular dating, 353–354, 362–369, 381, **449**
molecular evolution
gene duplication, 151–157
general principles, 86–89, 100
models of, 217–231, 242–244, 255
rate variation, 56, 115, 205–206, 363, 366
molecular presence/absence data, 391–392
molecular structural data, 392–395
monkeys
dispersal to New World, 5–6, 79
misconceptions of evolution, 30
relationships, 309–310
monophyletic group
as taxa. *See* monophyletic taxa
association with synapomorphies, 85–86
contrast to non-monophyletic group, 44–45, 122–124, 130, 136
definition of, 44–45, 121, **449**
exclusivity of, 122–123
named examples, 122–124, 174, 200, 411
species as, 163–165
See also clade
See also monophyly
monophyletic taxa, 123, 129, 165
monophyletic taxonomy, 126
monophyly
association with synapomorphy, 85–86
definition of, 44, 121, **449**
in taxonomy, 120–126, 165
of named taxa, 61, 174
of species, 165–166
See also exclusivity
See also monophyletic group
monotypic, 129, **449**
morning glory, 170
morphological data
carnivoran example, 174–181, 192–194, 200–201, 232–233, 244–245, 255–258, 270–271, 288
comparison among partitions, 285, 288, 291–292
comparison to molecular data, 88–89, 200–201, 257–258, 271, 280, 288, 291, 399
distance analysis, 232–233
encoding and scoring, 399–403

morphological data, *continued*
model-based approaches, 207, 230, 244
parsimony analysis, 192–194, 280
phylogenetic signal in, 22, 270–271
morphological traits
evolution of, 39, 83, 115
for phylogenetic analysis, 389, 395
in species concepts, 165
in taxonomy, 122, 130
See also morphological data
morphology. *See* morphological data; morphological traits
morphotypic data, 395–396, **449**
most parsimonious reconstruction, 97, 104, 184
most parsimonious trees
bootstrap, 275, 277, 301
carnivorans, 193–194, 200–201
partition homogeneity, 286–288, 299, 302
in phylogenetic inference, 95, 181–194, 200–205, 212–215, 297, 414, 416
searching for, 190–192, 416
mouse. *See* mice
MrBayes, 255, 257
multifurcation. *See* polytomy
multiple alignment, 199
See also sequence alignment
multiple changes, 82, 234
See also hidden changes
multiple data sets, 266, 285
multiple fossils, for molecular dating, 369
multiple hits. *See* hidden changes
multiple optima, 192
multiple regression, with PGLS, 339
museum specimens, 132, 166
mutant allele, 77, 80, 409
mutation rate, 221, 224, 226–227, 366
mutational bias, 320
mutations
branch length, 74
different kinds of, 196, 204, 393–395
different rates of gain and loss, 320–321
effect on visible phenotype, 77, 115
following lineage splitting, 39, 151
in Markov model, 218–219, 225–228
in trait evolution, 77–79, 86–87
independence of, 140
reversibility of, 179, 329
source of variation, 143
mutualists, 355–356, 358
Myotis, 294
naming clades, 132–133
naming taxa, 127–132, 165
narrative
adaptive, 315, 345
historical, 305–307, 422, **446**
key innovation, 375
linear, 28
natural classification, 119–120
natural group, 27
natural logarithms, 221, 239, 241
natural selection
as a driver of evolution, 78
in Darwin's theories, 12, 14–15, 23, 406
in the concept of adaptation, 306
See also selection
nectar spur
example of homology, 90–93
example of key innovation, 377–378
neighbor-joining (NJ)
main features, 234, 260, 275, **449**
application to carnivorans, 236
comparison to other method, 236, 238, 262
limitations, 268, 418
nested hierarchy, 20–21, 120, 128
nested models, 229, 246–247, 323, **449**
nested taxonomy (hierarchical classification), 20–21, 117–120, 126, 128, 164, 266, 268, **449**
nests, 85, 116
net diversification rate. *See* diversification rate
net-like, 40, 142, 157
See also reticulate
neutral evolution, 329
New World monkeys
dispersal history, 5, 79
relationships of, 309–310
Newick format, 47, 126, 128
niche conservatism, 360
niches, 359–360, 376, 396
NJ. *See* neighbor-joining
nodal (node) age, 54–55, 354, 365, 369
nodal depth/height, 362–368, **450**
node counting, 112, 113
node-based clade definition, 132, 136
nodes
absolute ages of, 354, 362–370, 386–387
ancestral states of, 315–319, 335–338, 342
considering all possible states for, 243
definition, 44, 68, **450**
effect of gene duplication, 152–156
for determining relatedness, 111–114
in clade naming, 126–127
in species trees, 292
in tree balance, 372–373
polytomous, 58
reconstructing biogeography, 353–354
relationship to trait evolution, 82, 100
relative ages of, 47, 53–55
rotation, 46
unrooted trees, 64
nomenclature, 107, 129–133, 138, 412, **450**
non-monophyletic, 44–45, 55, 85, 121–126, 411
nonclock models, 364, 366
nonoverlapping, taxa, 16, 266
nonparametric bootstrap, 273–281, 297, 301, **450**
normal distribution, 330, 366, 368
Nothofagus, 355
nuclear genes, 391
nucleotide position, 86, 90, 140, 143, 145–146, 195–196, 218, 221, 224, 230
nuisance parameters, 325, 417, 419, **450**
null hypothesis, 311–312, 323
O'Hara, Robert, 28, 35, 306
Occam's (or Ockham's) razor, 96

ocelot, 351
Ockham, William of, 96
offspring, 36, 143
Old World monkeys, 309–310
On the Origin of Species (Darwin, C.), 12, 14–15, 23
Onycophora, 371
order, taxonomic rank, 128, 130, 136
opossum, 132
opposable thumb, 307–309
optimal branch lengths, 237
optimal tree, 184, 189–190, 209, 242, 266, 273, 275, 287–288, 364
optimality criteria, 181, 236, **450**
optimality methods, 179
optimizing parameters, 244
optimum value of a trait, 336
orangutan, 295, 309–310
ordering changes on a branch, 82
ordering of tips, 46, 111–112
organellar genomes, 391
organismal pedigree, 140, 142, 148, 161, 164
Origin of Species. See *On the Origin of Species*
orioles, 335
Ornstein-Uhlenbeck (OU) model, 336, 339, 341, **450**
orthologs. *See* orthologous genes
orthologous genes, 152–153, **450**
ostrich, 167, 354–355, 394
otter. *See* carnivorans, worked example
OU model. *See* Ornstein-Uhlenbeck model
outgroup
  choice of, 175
  concept of, 64, 175–179, 182, **450**
  ultrametric trees, 362
outgroup method, 179
overall likelihood, 241, 243, 244
overall similarity
  contrast to relatedness, 416
  in species definition, 163
  See also similarity
overdispersion, 360–362
overly complex models, 247
overly simple models, 247
Pacific Ocean, 351
Pagel test, 323, 325, 347
paired phylogenies, 357–358
pairwise alignment, 199
pairwise distance, 232–234, 260, **450**
parallel evolution, 181, 392, **450**
paralogous genes (paralogs), 153, **450**
parameter space, 251–252
parameter-rich model, 246–247
parameters
  in maximum likelihood, 243–244, 246–247
  in the BiSSE model, 380
  in the Ornstein-Uhlenbeck model, 336
  in trait evolution, 324–325
parametric bootstrap
  for nodal depths, 365
  general features, 265, 281–283, 420, **451**
paraphyletic groups, 124–125, 136, 411, **451**
paraphyly. *See* paraphyletic groups
parasites, 355–358
parental DNA strand, 86
parenthetical tree format, 47, 126
parsimony
  bootstrapping with, 275, 277, 280
  clade support, 271–273, 300
  comparison to algorithmic methods, 236
  generalized or weighted. *See* generalized parsimony
  comparison to Hennigian inference, 181, 215, 416
  comparison to model-based methods, 217–218, 237–238, 250, 284, 318–319, 321, 326, 378, 422 312
  for phylogenetic inference, 173–237, 215, 282, 392–393
  for studying biogeography, 350–352, 355, 382
  for studying coevolution, 356–357, 385
  for studying gene duplication, 154–155, 171, 313–316
  for studying trait evolution, 95–99, 105, 306–310, 340, 347, 409
  in partition homogeneity test, 288, 302
  justification for, 185–187, 417
  phylogenetic signal, 268, 299
  principle of, 96–97, 100, 181, 186, 416–417, **453**
  problems of, 205–207
  species tree methods, 292
  squared change, 335, **455**
  See also maximum parsimony
parsimony cost, 205, **451**
parsimony-informative, 185, 209, 214, 266, **451**
parsimony-uninformative, 185, 193, 209, 414, **451**
partition, 230, 285–292, 302, 420, **451**
partition homogeneity test, 288–291, 302, 420, **451**
Pasteur, Louis, 10
paternal allele, 141, 143
path length, 231
PAUP*, 193, 200
peaks, in tree space, 191–192
pedigrees, 35–36, 107–109, 139, 166
penguins, 19
permutation tail probability test, 95, 268–270, 288, 299–300, **451**
permuted data sets, 268–269, 300
petals, 77, 79
PGLS, 339, **451**
phenetic species concept, 163, **451**
phenotype, contrasted to genotype, 77
phenotypic evolution, 79
Philosophie Zoologique (Lamarck), 10
phyletic series, 23–24
PhyloCode, 132–133, 138, **451**
phyloecology. *See* phylogenetic community ecology
phylogenetic clustering, 359–360, 384, 423, **451**
phylogenetic community ecology, 359–362, **451**
phylogenetic comparative methods, 337
phylogenetic correlation, 315, 361
phylogenetic evenness, 359
phylogenetic generalized least squares, 339, **451**
phylogenetic independent contrasts, 337–339, **452**
phylogenetic inference, 173–207, **452**
phylogenetic niche conservatism, 360, **452**

phylogenetic nomenclature, 107, 131–132, 138, 412, **452**
phylogenetic overdispersion, 359–361, 384, 423, **452**
phylogenetic reconstruction, See phylogenetic inference.
phylogenetic signal
indicators of, 266, 268–270, 291, 299
more in larger data set, 286
test for, 266–271
phylogenetic species concepts, 164–165, **452**
phylogenetic structure, 267–268, 339
Phylogenetic Systematics (Hennig) 27, 178
phylogenetic systematics, discipline of, 23, 25, 27, 85, 121, 130, 405–406, **452**
phylogenetic trees
as tools, 28
definition of, 1–2, 35, 40
importance of, 2
phylogenetic uncertainty, 58–59
phylogenetically randomized data, 266
phylogenetics. *See* phylogenetic systematics
phylogeny. *See* phylogenetic tree
phylograms, 56–57, 60, 74, 115, 163, 231, 378, **452**
Phylostrat, 80
phylum, 128, 130, 135
physiological traits, 79, 83, 163
pie charts, 317–318
pigment production, 77
pinnipeds, 174, 193–194, 200–201
Pisces, 121, 123–124
placental mammals, 277–278
plant phylogeny, 137, 394, 410
plastid DNA, 390–391
plastid genome, 56
platypus, 132
plesiomorphies, 85, 101, 121, 124–125, 179, **452**
point estimate
of a node's trait value, 336, **452**
of a tree, 242, 284, **452**
polar bears, 165
polarity of character change, 61, 84–85, 175, **452**
pollination, 323–324, 356, 377, 422
polymorphism, 177, 306, 397, **452**
polyphyletic groups, 124–125, 133, **452**
polyphyly. *See* polyphyletic groups
polyploidy, 151, **452**
polytomy, 58–59, 61, 74, 193, 258, **453**
population genetics, 293
population history, 139
population isolation, 38
population lineage, 147, 152, 156–158
population lineage splitting, 39, 156
See also lineage splitting
population reticulation, 156, 159, 297
population tree, 139, 146–148, 160–161, 171, 292
post-burnin sample, 255, 367
posterior distribution. *See* posterior probability distribution
posterior probability
accounting for different priors, 285, 420
accuracy of estimating, 284
estimation with MCMC 250–258, 418
in general, 247–250, 261, 417–418, **453**
of clades, 60, 303
of nodal depth, 367–368
of tree topology, 255–256
species trees, 294, 304
stochastic mapping, 325
posterior probability distribution
in general, 255–253, 284, 325–327, **453**
of ultrametric trees, 364, 368
posterior sample, 254, 312, 325, 365
predictive value of taxonomy, 4
presence/absence data, 390
primates, 71, 150, 205, 307–308
primitive, 23, 28
prior beliefs, 248, 258
prior distribution. *See* prior probability distribution
prior hypothesis, 284
prior probability
for rate variation, 366
in general, 248–249, 253–254, 262, 285, 418, **453**
of tree topology, 250
prior probability distribution, 258, 285, 325
priority at rank, 130, **453**
probabilistic model of evolution, 316, 323
See also mathematical models
products of evolution, species as, 162, 164
propinquity of descent, 120
proposal ratio, 252
proposal, in MCMC, 251–253
protein evolution, 231
protein sequence data, 389–390
proximity of tips, 111
pruning taxa, 49–51, 112, **453**
pseudoreplicate, 274–275, 281, 365, 420, **453**
PTP test. *See* permutation tail probability test
Punnett square, 226
pure birth model, 371–374
purines, 204, 228, **453**
pyrimidines, 204, 228, **453**
quadrupedal locomotion, 116
raccoons. *See* carnivorans, worked example
radial flower symmetry, 321–322, 388
radioactive decay, analogy to, 219
Rafflesia, 303
random data, 268, 270, 298
random partitions, 288, 290–291
random walk, 330
random-number generator, 281
randomized data, 266, 269
randomly amplified polymorphic DNA, 392
range expansion, 355
rank-based classification., 133
rank-free classification, 131
ranks (taxonomic), 128–131, 164–165, 412, **453**
RAPDs, 392
rapid evolution, as cause of no signal, 266
rapid morphological evolution, 115
rate categories, 230
rate heterogeneity, 227–230
rate homogeneity, 229
rate multiplier, in models of evolution, 228
rate of card evolution, 220
rate of change. *See* rate of evolution

rate of diversification, 371, 375–377, 378
rate of evolution
affects homoplasy, 203
effect on parsimony, 186
of continuous traits, 332
of discrete traits, 186, 219–220, 318, 322
of different branches, 56, 163, 207, 241, 363, 366–367
variation across genomes, 391
variation among sites, 186–187, 225, 230
rate of extinction, 370–372, 375–376, 379
rate of gains, 321
rate of lineage splitting, 370–371, 379
See also lineage diversification rate
rate of losses, 321
rate of molecular evolution, 56, 115, 230, 366
rate of sequence divergence, 224–225
rate of substitution. *See* rate of molecular evolution
ratite birds, 353–355
rats, 278
raw likelihood, 272
ray-finned fish, 28
reading across the tips, 111
reading frame, in a gene, 200
reading relationships from trees, 110
realizations of character history, in stochastic mapping, 325–326, 346
rearrangements, genetic, 190
recency of common ancestry
concept of, 25, 110
continuous traits, 334
contrast to similarity, 335
recombination
as cause of discordance, 143–146, 151
as cause of indels, 197
recombination rate, 145
recombinational gene, 146, 151, 161, **453**
rectangular tree format, 43, 48, 58, 73, **453**
rectangular-right format, 48
relatedness, 4, 42, 107–117, 133, **453**
relative ages of nodes, 53–54
relative chronograms, 362, 364–365, 367, 369, 380, **454**
relative likelihood, 317–319
relative nodal age, 365
See also nodal age
relative posterior probability, 250, 325
relative rate parameter, 228
relaxed molecular clock, 363, 366–367, **454**
reproduction within populations, 35
reproductive isolation, 39, 160–161, 164, 413, **454**
reptiles, 27, 117, 120–124
resolved (trees or clades), 58–59, **454**
restriction fragment length polymorphisms
as frequency data, 396
for phylogenetic inference, 391–392
models for, 231
restriction mapping, 393
retention index, 94
reticulate evolution. *See* reticulation
reticulation, 157, 160–161, 294, 296, **454**
reversal
in general, 179, 183, **454**
of a morphological transition, 80, 92–93
reverse mutation, 78
RFLP. *See* restriction fragment length polymorphisms
RI. *See* retention index
RNA sequences, 390
root, 44, 61–63, 65, 175, **454**
rooted trees, contrast to unrooted, 53, 61, 64, 214, **454**
rooting
outgroup method, 61, 64, 175
principles, 64, **454**
under clock models, 64, 362
rotating branches, 112
rotating nodes, 46
saddle joint, 307
salamander, 102, 110–114, 124
sampling with replacement, 274, 281, **454**
SARS, 6
Scaevola, 167
Scala Naturae, 10, 406
scales, 80, 85
scapula. *See* dorsal scapula
scientific hypotheses, trees as, 265
scoring, of morphological traits, 176–177, 181, 300–403, **441**
sea lions. *See* carnivorans, worked example
seals. *See* carnivorans, worked example
second world war, 305–307
secondary biochemistry, 395
section, taxonomic rank, 128
segmental duplication, 151, **454**
selection
effect on rate of evolution, 230
effect on rate of fixation, 79
favored opposable thumb, 309
in adaptation, 306, 349
selective advantage, testing for, 307
selective bias, 320
selective regime, 308–309, 311–312, **454**
self-compatibilty, 380–381
self-incompatibility, 380–381
separate ancestry, 10, 15, 31–32
Seq-Gen, 281
sequence alignment
by eye, 199–200, 207, **454**
principles, 89, 178, 195–200
programs, 199
sequence evolution, 86–88
See also molecular evolution
sequence variants, 141
series, taxonomic rank, 128
Severe Acute Respiratory Syndrome, 6
sex, effect on diversification, 388
sexual dimorphism, 341
sexual populations, 140–143, 146
sexual reproduction, 141, 160
sexual selection, 319
shared ancestral character states, 85, 124
shared clades, among trees, 66
shared derived characters, 85, 416
See also synapomorphies
shiitake mushrooms, 364–365
shorter, regarding trees, 187, 268, **454**
shoulder, evolution of, 307–308, 310
showy flowers, 323–325

shuffling, 268
See also permuted data
sibling, 108
signal, phylogenetic, 186, 265
signposts method for reading trees, 46–48
silent substitutions, 86, 115, **454**
simian primates, 308–310
similarity
compared to relatedness, 410
contrast to common ancestry, 25, 27, 335, 416
contrast to relatedness, 108, 114–117, 406
evidence for common ancestry, 15–19, 91
gene duplication, 154
species concepts, 163
simulating
character histories in stochastic mapping, 325–327
data, 273, 281
molecular evolution, 266
phylogenetically random data, 21, 268
sister clade, 372–373
sister group, 45, 338, **454**
sister lineage, 56
sister taxon. *See* taxon, sister
site likelihood, 241, 243–244, 259, **454**
skin pigmentation, 340
skin reflectance, 339
skulls, 174, 340
slowly evolving characters, 187
slowly evolving DNA regions, 203
snakes
accelerated diversification, 375
biogeography example, 382
limb loss, 80, 136
relationships of, 27, 102, 123–124, 340
soft polytomies, 59, **455**
Solanaceae, 380–381
South Africa, 360
South America, 351, 353–355
special creation, 15
speciation
concept of, 38, 357
causes of, 39
See also lineage splitting
species
as clades, 164–165
as players in evolution, 162, 164
as a rank, 128, 165
as tips, 43
concept of, 139, 161–166
co-occurrence, 359, 423
concepts of, 139, 162
description in taxonomy, 165
relationships, similarity to familial, 110
species richness, 371
species tree, 156–157, 171, 292–294, 304, 420–421, **455**
See also population tree
species tree inference
in general, 292–294
molecular presence/absence data, 392
specific epithet, 128, **455**
specifiers, 132–133, 136, 412
spiders, 11, 313–315, 371
splits, 47, 61, 63, **455**
splits format, 46
splitting, of population lineages. *See* lineage splitting
spontaneous generation, 9–10
SPR. *See* subtree pruning and regrafting
squared-change parsimony, 335, **455**
squirrel, 50, 79, 186
stabilizing selection, 336
standardized contrast, 339
standardized endings, ranks, 130
star phylogeny, 266
state-dependent diversification. *See* character-state dependent diversification
stationarity, 254, **455**
statistical inference, role in phylogenetics, 265, 349
statistical methods (statistical analyses)
ancestral state reconstruction, 315
evolutionary correlation, 311–312, 327, 338, 379, 422
for phylogenetic inference, 28
need for, 187, 286
statistical power, 217, 237, 247, 374, **455**
statistical tests
of evidence for evolution, 21–23
of cophylogeny, 358
in phylogenetic inference, 173, 186–187, 265–297, 403
STEM, 294
stem lineage, 84, 318, **455**
stem node, 131, **455**
step matrix, 204–205, 315
steps, in parsimony, 182, 270, **455**
stochastic mapping, 325–327, 346–347, 421, **455**
storing information, using phylogenies, 84
strata, fossil-bearing, 40
Strepsiptera, 282–285
strict consensus, 193–194, **455**
structural molecular data, 394
studying trait evolution, 323
subclades, 127, 200, 373–375
subfamily, taxonomic rank, 130
subjectivity, of Bayesian prior, 258
subspecies, taxonomic rank, 128, 165
substitution matrix. *See* substitution probability matrix
substitution models, 218–229, **455**
substitution probabilities, 227, 316–317, 325
substitution probability matrix, 219–224, 227, 259, 317, **456**
substitutions
expected number of, 219–220, 226–227
models of, 218–231, 241, 259, 263–265, 418
molecular evolution, 196–200, 390
rate of, 219–222
types of, 228–229
subtree, 66
subtree pruning and regrafting (SPR), 66–67, 75, 190, 408, **456**
suit, as metaphor for base, 219–220
sum of the squared deviations, 237
sum of tree lengths, in partition homogeneity test, 288–291
summarizing a posterior distribution, 255
summing over histories, 244

sunfish, 361
support, for clades, 60–61, 271–274, 278, 281, 294, 297, 403,
swordtails, 319, 326
symmetric model, 317, 320–322
sympatric, 39, **456**
symplesiomorphy, 85, 116, **456**
synapomorphy, 85–86, 124, 394, **456**
synonymous substitutions. *See* silent substitutions
systematics, discipline of, 23–27, 30, 107, **456**
systematization, 123
  See also classification
tails, 116, 175–176
tapeworms, 356–357
taxon (taxa)
  as tips/terminals, 3–5, 44, 48–52, 61–64, 69–70, 75, 137, 212–215
  definition of, 4, 44, **456**
  in data matrices, 94, 174–178, 182–190, 208–209, 212–215, 301
  monophyly of, 52, 85, 121–128, 164–165
  naming, 129–133
  number included, 67, 174, 112, 237, 243, 283, 373, 380
  sister, 45, 53, 182, 201, 285, **454**
  See also taxonomy
  See also terminals; tips
taxonomic rank, 107, 128–133, 136, 164–165, 412, **453**
taxonomic system. *See* taxonomy
taxonomies, converting into trees, 126–128
taxonomy
  definition of, 107, 117, **456**
  history of, 25, 28, 117–24,
  impact of phylogenetic systematics, 28, 107, 117–133, 405
  vertebrate, 117–124
  See also classification
teeth, 116, 174, 181, 204
Templeton test, 273, 287–288, 292, **456**
terminal branch, 115, 148, 157, 163, 185, 309, **456**
terminal, 44, 49, 52, 115
  See also tip
testing hypotheses of adaptation, 306–311
testing phylogenetic hypotheses, 265–304
tetrapods, 28, 117, 123–125, 136
the comparative method. *See* comparative method
thermoregulation, 307
third codon positions, 203, 386
threshold of significance, 60, 240
throwing ability, 308–310
thumbs, 308–309
tigers, 186, 351
Tiktaalik, 114
time axis, 53–55, 73
time calibrate. *See* calibration
time reversibility, 218, 226–228, **456**
time slice, 41
timescale, 37–38, 54
tip, 35, 41, 44, 175, **456**
tip correlation, 312, 317, 329, **456**
tip data, 317, 339
tip order, 112–113
tomatoes, 295–296
tool use, in primates, 307–309
topology test, 273, 285, 287–288, **456**
topology. *See* tree topology
total branch length, 237
total evidence. *See* combined analysis
trait correlation. *See* correlated evolution
trait value, expected under Brownian motion, 332, 335
trait-based species concepts, 162–165, **457**
transition matrices, for amino acids, 390
transition probability matrix. *See* substitution probability matrix
transition:transversion bias, 228–229, 246, 252, **457**
transitional fossils, 16, 19, 33, 405, **457**
transitions, 204–205, 228–229
translocations, 196, 393
transthyretin 2 gene, 178, 200, 210
transversions, 204–205, 228–229, **457**
tree, definition of, 2, **457**
tree balance, 372, **457**
tree building, 265
tree formats. *See* tree styles
tree length
  in parsimony, 182–190, 209, 213, 236, 414, **457**
  in stochastic mapping, 325
  measure of phylogenetic signal, 269–272, 290–291
tree metaphor, 11
tree of life
  in history of systematics, 23, 25–26, 107
  knowledge of, 133
  metaphor, 14, 17, 20, 23, 139, 164
  organizing knowledge of, 3, 7, 83–84, 311
  single common ancestor, 41, 49
  size of, 49
  variation in rate, 114, 371
tree quality, 190, 236, 240
tree rearrangement, 66–67, 190–191
tree reconstruction. *See* phylogenetic inference
tree space
  in Bayesian MCMC, 285, 418, **457**
  in relation to clade support, 272–273
  searching for trees in, 67, 190–192, 251, 255
tree styles, 48
tree thinking
  as a facet of biological literacy, 7
  characteristics of, 1, 9, 15, 33, 35, 41, 161, 406
  compared to ladder thinking, 29–30, 33
  definition of, 1, **457**
  despite reticulation, 160
  importance of, 1, 2, 7
  philosophy of history, 306
tree topology
  compatibility of two trees, 50–52
  definition of, 45–46
  effect of diversification rate, 374
  equivalence of two trees, 45–48, 62–65
  in Bayesian inference, 250, 252, 255–257, 284–285
  in minimum evolution, 237

tree topology, *continued*
in maximum likelihood, 240, 281
in parsimony, 207
information content, 45–48, 50–55, 67, 112, 250
prior probability of, 250, 287
tree space, 272–273
tree-to-tree distance, 65–67
unrooted, 62–65
tree-o-scope, 145
tree-to-tree distance, 65, **457**
trend, 320, **457**
tribe, taxonomic rank, 128, 130
turtles, 123–124, 134
type specimen, 130, **457**
ultrametric trees
for BiSSE, 380
for molecular dating, 362, 364, 366–367, **457**
unbalanced trees, 373
uncertainty
in character scoring, 176, 185
in nodal depth, 365
in topology, 312
of ancestral states, 312, 319, 327, 332, 338
of dispersal history, 352
portraying with consensus, 59
unconstrained search, 300
uncorrelated relaxed molecular clock, 366, **458**
unequal base frequencies, 228–229, 242
unequal rates of evolution, 115
unequal rates of gain and loss, 379
ungulates
independent contrasts, 338–339
phylogeny of, 347, 422
tapeworm hosts, 356
uniform prior. *See* flat prior probability
uninformative, in parsimony, 214, 415
uninformative prior probability. *See* flat prior probability
uniparental inheritance, 140
unknown, character-state scoring, 176, 185
unlinked loci, 143
unrooted trees, 61–65, 67, 414, **458**
uphill proposals, 251
using trees to organize data, 4–5, 83–84
variance
in bootstrap replicate, 275
in clade size, 375
in trait value, 330–332
under Ornstein-Uhlenbeck model, 336
variety, taxonomic rank, 128
Venn diagrams, 118–120, 126–128, 137
vertebral column, 118, 120–123
vertebrates
classification and evolution, 71, 117–124
diversification rate variation, 374–375
evolutionary advancement of, 28–29
vestigial traits, 16, 31, **458**
vicariance, 352–356, 362, 383, **458**
viruses
causes of lateral gene transfer, 158
phylogenetic trees of, 41, 205
SARS, 6
waiting time
in pure birth model, 371–372
in substitution model, 219, 227
walking
evolution of, 28
through tree space, 192, 251–252
Wallace, Alfred Russel, 9, 350
walruses. *See* carnivorans, worked example
warm blooded animals, 11–12, 124
wasps
fig pollinators, 356–358
gall, 318
weighted parsimony. *See* generalized parsimony
weighting. *See* character-state weighting or generalized parsimony
whales
blowhole homology, 90
evolutionary narrative for, 422
features of, 16, 80, 347, 422
relationships, 134–135, 278–279, 347
whole genome duplication, 151, **458**
wind pollination, 323–324
wings
evolution in insects, 95–98
example of convergence, 3, 31, 90
vertebrate classification, 117–118, 120–122
Zimmerman, Walter, 25, 32
Zoological code, 130
Zoonomia (Darwin, E.), 11–12
zygomorphy (bilateral flower symmetry)
example of evolutionary trend, 321–322
key innovation example, 388